In vivo characterization of RNA cis-regulators in bacteria

Author: Arianne M. Babina

In vivo characterization of RNA *cis*-regulators in bacteria

Arianne M. Babina

A Dissertation

submitted to the Faculty of

the Department of Biology

in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Boston College
Morrissey College of Arts and Sciences
Graduate School

December 2017

Abstract

In vivo characterization of RNA *cis*-regulators in bacteria

Arianne M. Babina

Advisor: Michelle M. Meyer, Ph.D.

Bacteria commonly utilize *cis*-acting mRNA structures that bind specific molecules to control gene expression in response to changing cellular conditions. Examples of these ligand-sensing RNA *cis*-regulators are found throughout the bacterial world and include riboswitches, which interact with small metabolites to modulate the expression of fundamental metabolic genes, and the RNA structures that bind select ribosomal proteins to regulate entire ribosomal protein operons. Despite advances in both non-coding RNA discovery and validation, many predicted regulatory RNA motifs remain uncharacterized and little work has examined how RNA *cis*-regulators behave within their physiological context in the cell. Furthermore, it is not well understood how structured RNA regulators emerge and are maintained within bacterial genomes. In this thesis, I validate the biological function of a conserved RNA *cis*-regulator of ribosomal protein synthesis previously discovered by my group using bioinformatic approaches. I then investigate how bacteria respond to the loss of two different *cis*-regulatory RNA structures. Using *Bacillus subtilis* as a model organism, I introduce point mutations into the native loci of the ribosomal protein L20-interacting RNA *cis*-regulator and the tandem glycine riboswitch and assay the strains for fitness defects. I find that disrupting these regulatory RNA structures results in severe mutant phenotypes, especially under harsh conditions such as low temperatures or high glycine concentrations. Together, this body of work highlights the advantages of examining RNA behavior within its biological context and emphasizes the important role RNA *cis*-regulators play in overall organismal

viability. My studies shed light on the selective pressures that impact structured RNA evolution *in vivo* and reinforce the potential of *cis*-regulatory RNAs as novel antimicrobial targets.

Dedication

To Spike,

This was all for you.

—

Awake, arise, or be for ever fall'n!

Paradise Lost, John Milton

—

Do I dare

Disturb the universe?

In a minute there is time

For decisions and revisions which a minute will reverse.

The Love Song of J. Alfred Prufrock, T. S. Eliot

Acknowledgements

To my advisor, Michelle Meyer: Words cannot adequately convey the utmost gratitude, appreciation, and admiration that I have for all of the knowledge, personal growth, guidance, patience, and encouragement you have given me over the past six (plus) years. You are one of the most enthusiastic, driven, and hard working individuals I have ever encountered and your passion for your field of research is inspiring, motivating, and contagious. Thank you for making me the scientist that I am today.

To my seventh grade science teacher, Peter Bavone: Thank you for introducing me to biology and the wonderful world of single-celled organisms. I haven't stopped exploring since!

To Dieter Söll and Lennart Randau: Thank you for taking your chances on a young and naïve undergraduate so many years ago and for opening my eyes to the beauty of small RNAs in bacteria. I truly would not be where I am today without your continued support, kindness, and generosity. I am forever grateful for all of the wisdom you have imparted on me and I am honored to consider you both colleagues and friends.

To Stephane Bentolila and Maureen Hanson: Thank you for giving me the opportunity to branch out of my comfort zone and grow as a young scientist; I learned so much from both of you. Stephane, thank you for your tough love, honest feedback, (sometimes unsolicited) fatherly advice, teaching me how to take a step back and breathe, and for always pushing me to realize my full potential.

To my thesis committee, Tim van Opijnen, Welkin Johnson, Charlie Hoffman, and Babak Momeni: Thank you for the helpful advice, guidance, laughs, and words of encouragement. You always managed to transform my committee meetings into fun and stimulating scientific discussions. It was a pleasure to have you all by my side every step of the way. Thank you for always having your doors open, teaching me the value of cost-benefit analysis, and for the countless supplies I borrowed from your labs over the years.

To my lab mates, Yang Fu, Betty Slinger, and Shermin Pei: Thank you for the years of friendship, encouragement, coffee breaks, dinner dates, Thirsty Thursdays, and game nights. Thank you for your passion and the many hours spent troubleshooting, complaining, and dreaming. More recently, thank you for only being an email or text away. You are incredible friends, colleagues, and scientists, and I can't wait to see what great things you will accomplish in the future.

To the many undergraduate and rotation students that I have mentored, especially Nicholas Lea and Amalia Brawley: Thank you for teaching me patience, humility, and the art of troubleshooting. Thank you for your hard work, dedication, genuine interest, and above all else, your friendship. It was a privilege mentoring you and watching you grow. I am so proud of your accomplishments in the lab and of the amazing individuals you have become.

To my family, especially my parents, Christine and David Babina: Thank you for your unconditional love and support, for always thinking the best of me, for never questioning my chosen path, and most importantly, for never asking me "when are you going to graduate?"

To my husband, Adam Kirsch: Thank you for being my best friend and for always believing in me, especially when I don't always believe in myself. Thank you for giving me the freedom to pursue my dreams, for being able to see beyond the present, and for always looking forward to the future. Thank you for the countless rides to and from lab late at night and on weekends, the dinner deliveries, and the good company when time points made long nights turn into mornings.

Table of Contents

List of Figures

List of Tables

Abbreviations

ATP	adenosine triphosphate
bp	base pair
cpm	counts per minute
DIC	differential interference contrast (microscopy)
DNA	deoxyribonucleic acid
cDNA	complementary DNA
gDNA	genomic DNA
EDTA	ethylenediaminetetraacetic acid
EMSA	electrophoretic mobility shift assay
IPTG	isopropyl β-D-1-thiogalactopyranoside
kb	kilobase (pair)
MOPS	3-(*N*-morpholino)propanesulfonic acid
OD	optical density
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
qRT-PCR	quantitative RT-PCR
RACE	rapid amplification of cDNA ends
RBS	ribosome-binding site
RLM-RACE	RNA ligase-mediated RACE
RNA	ribonucleic acid
mRNA	messenger RNA
rRNA	ribosomal RNA
tRNA	transfer RNA

RT	reverse transcription
SD	Shine-Dalgarno (sequence)
SDS	sodium dodecyl sulfate
TE	tris-EDTA (buffer)
UTR	untranslated region
WT	wild type

Chapter 1

Introduction to RNA *cis*-regulators in bacteria

In bacteria, transcription and translation occur simultaneously (Proshkin et al. 2010; McGary and Nudler 2013). As a result, bacteria have evolved a number of ways to exploit the close linkage of these two important processes to efficiently control gene expression for rapid adaptation to changes within their environments. One manner in which bacteria regulate gene expression is through the use of messenger RNA (mRNA) *cis*-regulators, RNA motifs that are often structured and mediate the expression of the transcripts in which they are contained.

Structural features of bacterial mRNA transcripts play integral roles in modulating transcription and translation efficiency and mRNA degradation. During transcription, the nascent mRNA can base pair with itself to form secondary structures that induce pausing or termination of the actively transcribing RNA polymerase (Gusarov and Nudler 1999; Yarnell and Roberts 1999), prevent ribosome access and translation initiation, and/or stall concurrently translating ribosomes. Transcript conformations stabilized by

the presence of a paused RNA polymerase or ribosome can protect the mRNA from premature interactions with transcription termination factors (such as Rho) or nucleases. Alterations in mRNA folding and stability in response to changing cellular conditions such as temperature and pH can influence transcription, translation, and degradation rates and allow for condition-specific gene expression patterns (for review see: (Meyer 2017)).

Certain mRNA sequences can adopt complex three-dimensional shapes to modulate downstream expression in response to interactions with specific molecules. These structured RNA "sensors" are commonly employed by bacterial genomes to regulate fundamental processes, such as the production of ribosomal components and metabolic pathways essential for survival or pathogenesis. The best-characterized examples of bacterial RNA *cis*-regulators that respond to ligand-binding interactions are those that control ribosomal protein synthesis and riboswitches.

1.1 RNA *cis*-regulators of ribosomal protein synthesis

1.1.1 The prokaryotic ribosome

The ribosome is a highly complex molecular machine that plays an essential role in protein biosynthesis in all living organisms. The prokaryotic 70S ribosome consists of two subunits, each of which is composed of ribosomal RNA (rRNA) and many ribosomal proteins. The small 30S ribosomal subunit, which interacts with the mRNA to be translated, consists of the 16S rRNA and approximately 21 ribosomal proteins. The large 50S subunit, which participates in the polymerization of nascent polypeptide chains, is composed of the 5S and 23S rRNAs and about 33 different ribosomal proteins (for review see: (Kaczanowska and Rydén-Aulin 2007; Shajani et al. 2011)). In rapidly dividing *Escherichia coli*, as much as 40% of total cellular dry mass is comprised of

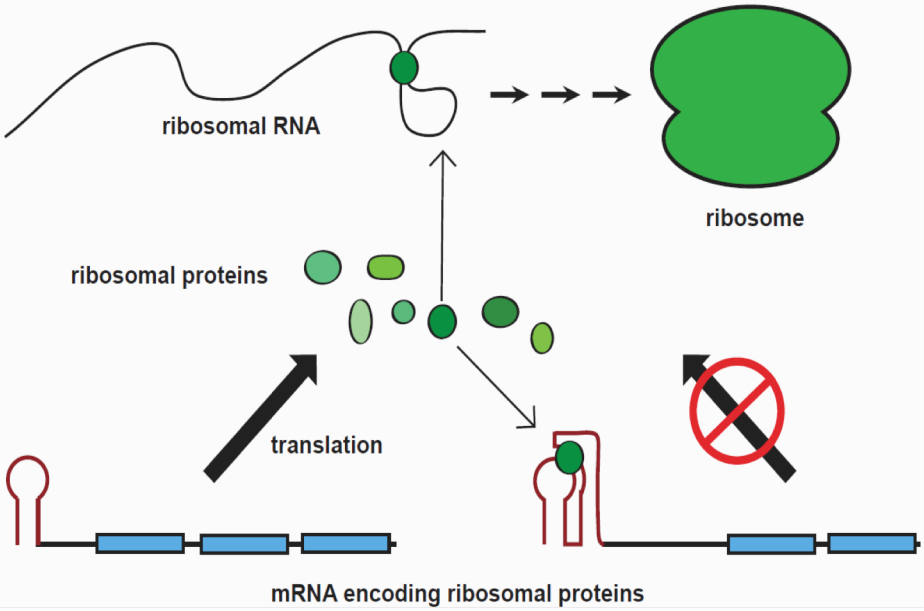


Figure 1.1. Autogenous regulation of ribosomal protein synthesis in bacteria. During ribosome assembly, ribosomal proteins typically bind to specific sites on rRNA. When rRNAs are saturated with bound proteins or when ribosomal proteins are in excess, select ribosomal proteins can interact with RNA structures located within their operon transcripts to inhibit further ribosomal protein expression at the transcriptional or translational level.

ribosomes and their associated cofactors (Tissieres et al. 1959; Bremer and Dennis 1996) and ribosome production consumes about 40% of the cell's total energy (Nierhaus 1991). Because of this enormous energy expense and the importance of appropriate component stoichiometry in ribosome biosynthesis and function, the production of individual ribosomal components and extra-ribosomal cofactors is highly regulated and tightly coordinated (Harvey 1970; Bremer and Dennis 1996).

1.1.2 Regulation of ribosomal protein synthesis in bacteria

One manner in which bacteria maintain the delicate balance between rRNA, ribosomal proteins, and extra-ribosomal cofactors to ensure proper and efficient ribosome production is through the autogenous regulation of ribosomal protein synthesis (Nomura et al. 1980, 1984; Zengel and Lindahl 1994). In bacterial genomes, the genes

coding for ribosomal proteins exist as single copies and are clustered together in several large polycistronic operons. When cells are rapidly dividing, rRNA and ribosomal proteins are produced *en masse* to satisfy the cells' increasing need for active protein synthesis. During ribosome assembly, ribosomal proteins preferentially bind to their target rRNAs. However, when rapid ribosome synthesis is no longer needed and/or rRNA-binding sites are saturated with bound proteins, select ribosomal proteins will bind to structured non-coding RNA motifs that are typically located in the 5'-untranslated regions (UTRs) of their own operon transcripts (Figure 1.1). These interactions are highly specific and induce conformational changes within the mRNA that repress transcript expression at the transcriptional or translational level. In most instances, the *cis*-regulatory RNA elements are molecular mimics of rRNA structure and their binding partners are primary rRNA-binding ribosomal proteins (Fu et al. 2013). While most of these RNA *cis*-regulators respond to a single effector protein, some require interactions with a complex of ribosomal proteins to carry out their regulatory function, such as the RNA structures that interact with the S6:S18 heterodimer (Matelska et al. 2013; Fu et al. 2014; Babina et al. 2015) or the pentameric L10$(L12)_4$ complex (Iben and Draper 2008; Yakhnin et al. 2015).

1.1.3 Diversity and distribution of ribosomal protein RNA *cis*-regulators

This RNA-based negative feedback regulation has been extensively characterized in the model Gram-negative bacterium, *E. coli*. Fifteen distinct RNA *cis*-regulatory elements have been found to control the synthesis of over half of the ribosomal proteins within this organism (Figure 1.2) (Zengel and Lindahl 1994; Fu et al. 2013; Matelska et al. 2013; Fu et al. 2014; Aseev et al. 2015, 2016). These *cis*-

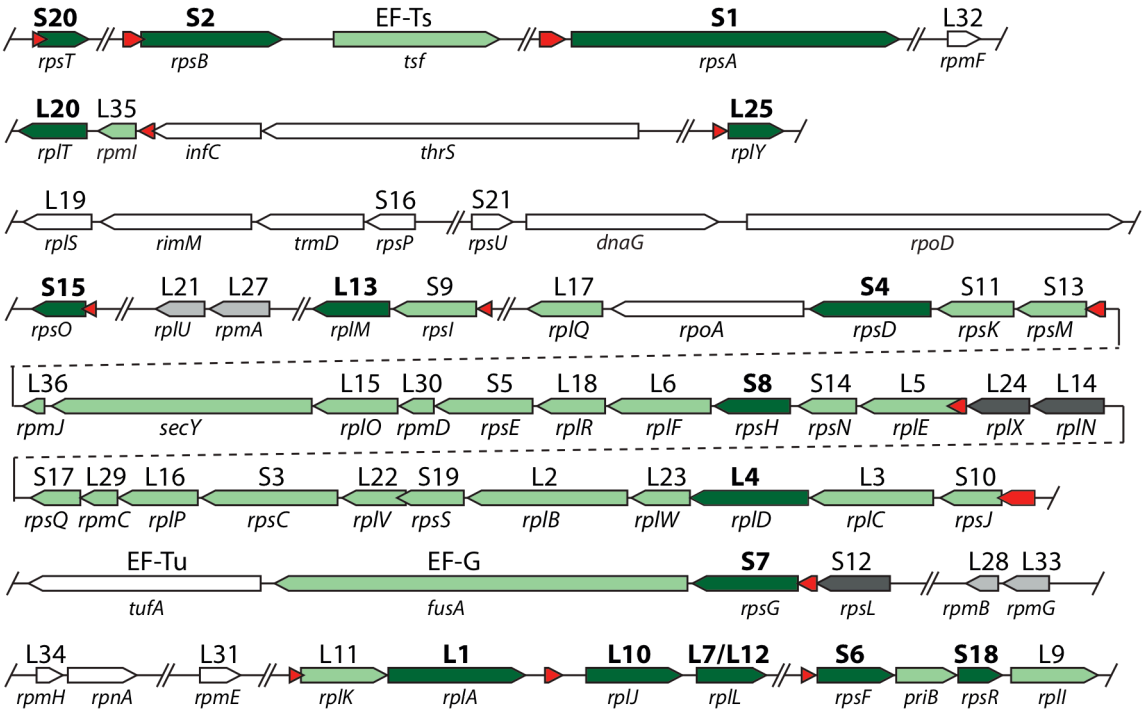


Figure 1.2. Ribosomal protein operon organization and regulation in *E. coli*. Gene names are given below each arrow and ribosomal protein product names are given above each arrow. Red arrows represent experimentally validated regulatory RNA structures; the two RNA structures that interact with ribosomal protein L20 are represented by a single arrow. Green arrows represent genes that are autogenously regulated; dark green arrows indicate the effector proteins responsible for regulation. Dark gray arrows are genes with reported autogenous regulation and white arrows indicate genes with no known autogenous regulation. Light gray arrows are confirmed genes not subject to autogenous regulation. Double slashes indicate breaks in genome sequence. Adapted from (Fu et al. 2013).

regulatory RNA structures exhibit a great deal of variation in size and complexity and utilize different mechanisms to inhibit gene expression.

For example, the S4-interacting and S15-interacting RNA regulatory elements found in *E. coli* contain complex pseudoknotted structures and inhibit translation of their corresponding operons via ribosomal entrapment (Jinks-Robertson and Nomura 1982; Thomas et al. 1987; Tang and Draper 1989; Portier et al. 1990; Philippe et al. 1990, 1993, 1995; Baker and Draper 1995). Conformational changes of the mRNA occurring upon protein binding trap the 30S ribosomal subunit, forming a ternary complex that impedes translation initiation. The S1-interacting RNA is proposed to control expression

of its own transcript using a similar mechanism, although its three-helix structure is considerably less complex (Boni and Artamonova 2000; Boni et al. 2001). The L1-interacting RNA consists of a simple stem-loop with an internal bulge that is a direct mimic of the L1-binding site on the 23S rRNA. L1 binding to the mRNA inhibits translation through a ribosome displacement mechanism in which the 30S subunit is blocked from accessing the ribosome-binding site on the transcript (Dean and Nomura 1980; Baughman and Nomura 1983). The large L4-interacting RNA regulatory element contains multiple helices and uniquely regulates expression of the operon it precedes using both translation inhibition and NusA-dependent transcription termination mechanisms (Yates and Nomura 1980; Lindahl et al. 1983; Freedman et al. 1987; Shen et al. 1988; Zengel and Lindahl 1992, 1994; Sha et al. 1995; Zengel and Lindahl 1996; Zengel et al. 2002).

Four of the ribosomal protein RNA *cis*-regulators found in *E. coli* (interacting with ribosomal proteins S2, S6:S18, L1, and L10(L12)$_4$), are widely distributed across most bacterial phyla (Figure 1.3). All bacterial phyla, except Acidobacteria, possess at least one of these *cis*-regulatory RNA structures (Fu et al. 2013; Matelska et al. 2013; Fu et al. 2014), setting the precedent that the autogenous regulation of ribosomal protein synthesis is nearly universal in bacteria. The wide distribution and conservation of these specific RNA elements suggests vertical inheritance from an ancient ancestor. This is further supported by the fact that examples of these conserved RNA structures are able to interact with corresponding ribosomal protein homologs from different bacterial species (Köhrer et al. 1998). The remaining eleven ribosomal protein RNA *cis*-regulators found in *E. coli* are relegated to only a few orders of Gammaproteobacteria (Fu et al. 2013).

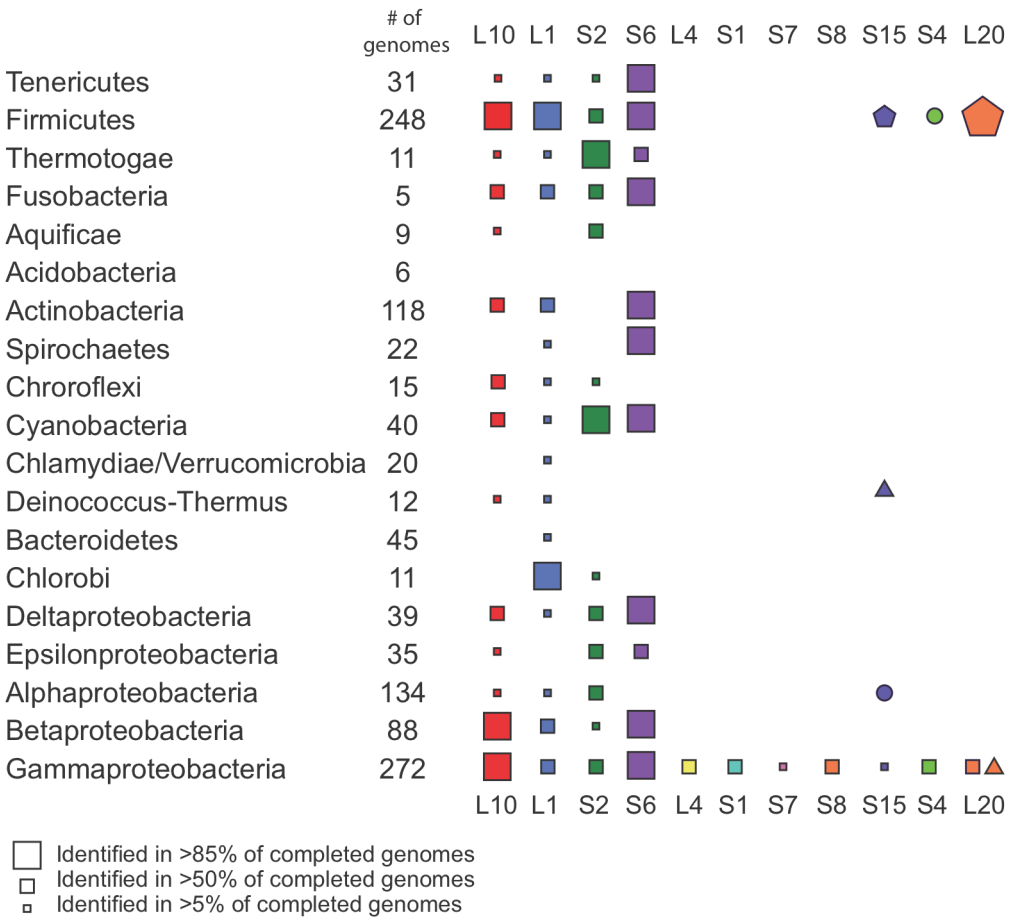


Figure 1.3. Distribution of ribosomal protein regulatory RNA structures in bacterial phyla. Distributions were compiled from alignments for twelve of the fifteen RNA structures controlling ribosomal protein biosynthesis in *E. coli*. Different shapes within the same column represent distinct RNA structures that interact with and/or precede the gene encoding the indicated ribosomal protein. Adapted from (Fu et al. 2013).

Recent structure-based homology searches and phylogenetic analyses have identified several independently derived RNA structures that perform analogous regulatory functions in different bacterial phyla (Deiorio-Haggar et al. 2013; Slinger et al. 2014). While some features that mimic rRNA protein-binding sites are conserved, these RNA regulators are narrowly distributed to select phyla and share little to no sequence and/or structural similarity to the regulatory RNAs found in *E. coli* that interact with homologous ribosomal proteins (Figure 1.3). The RNA structures that interact with ribosomal protein S15 are perhaps the most striking illustration of this diversity (Figure

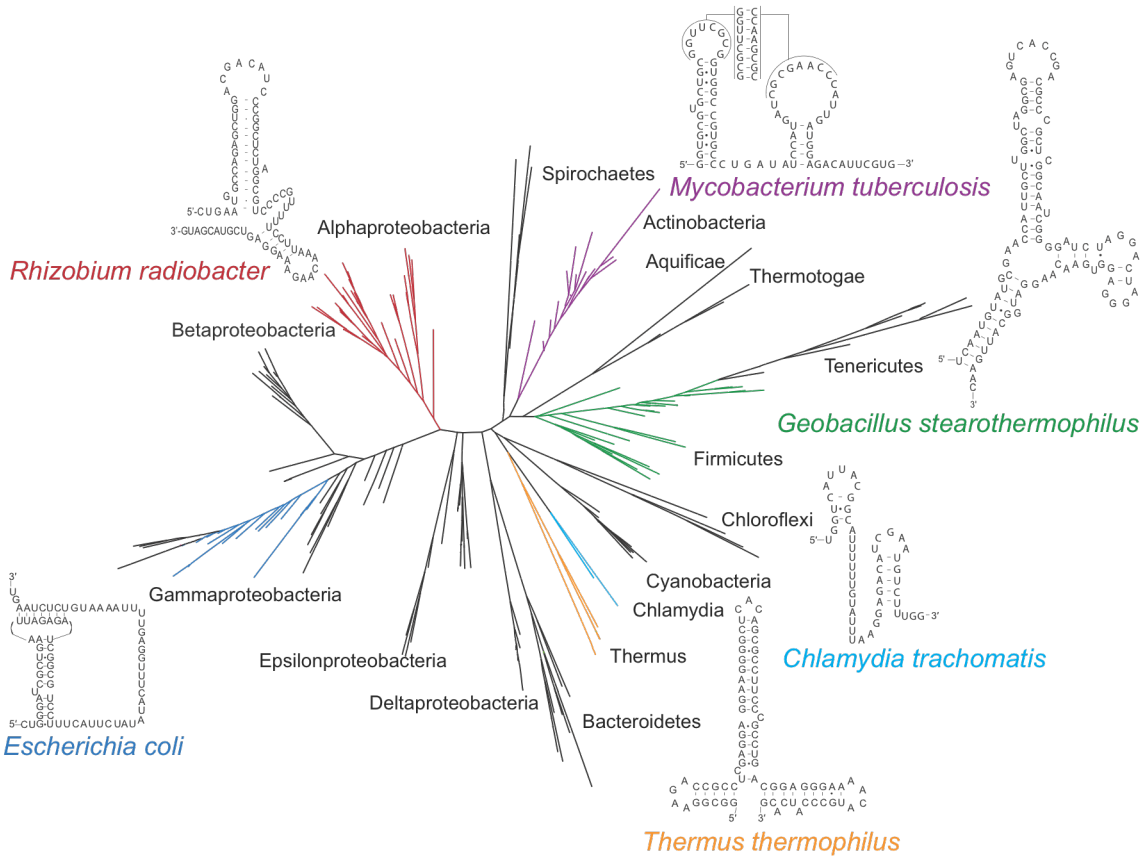


Figure 1.4. The diversity of the S15-interacting RNA *cis*-regulator. The RNA structures that regulate gene expression in response to ribosomal protein S15 are narrowly distributed to select bacterial phyla. The regulatory structures found in *E. coli*, *Rhizobium radiobacter*, *Geobacillus stearothermophilus*, and *Thermus thermophilus* have been experimentally verified; those found in *Mycobacterium tuberculosis* and *Chlamydia trachomatis* have been computationally predicted but have yet to be biologically validated (Slinger et al. 2014). Adapted from (Slinger 2016).

1.4). Across different groups of bacteria, four distinct S15-interacting regulatory RNAs have been experimentally validated and two additional structures have been predicted (Philippe et al. 1990; Scott and Williamson 2001; Serganov et al. 2003; Slinger et al. 2014; Slinger 2016).

More examples of these independently derived RNA *cis*-regulators can be found in the Gram-positive bacterium *Bacillus subtilis.* This organism contains the four widely distributed *cis*-regulatory RNA structures as well as three that are specific to the Bacilli class: the RNA structures interacting with ribosomal proteins L20, S4, and S15 (Fu et al.

2013; Matelska et al. 2013; Deiorio-Haggar et al. 2013; Fu et al. 2014). Unlike in *E. coli*, the *B. subtilis* S4-interacting RNA is a multi-helical structure that does not form a pseudoknot and represses protein synthesis shortly after transcription initiation (Grundy and Henkin 1991, 1992). The *B. subtilis* L20-interacting RNA is composed of three helices and controls the expression of its operon via the formation of a Rho-independent transcription terminator (Choonee et al. 2007). Conversely, there are two L20-interacting RNA *cis*-regulators in *E. coli* that regulate downstream gene expression at the translational level; one structure consists of a single bulged helix and the other contains a long range pseudoknot (Lesage et al. 1990, 1992, Guillier et al. 2005a, 2005b). Moreover, although the protein-binding site of the L10$(L12)_4$-interacting RNA is conserved between *E. coli* and *B. subtilis*, the regulatory mechanism of action is not (Fu et al. 2013). In *E. coli*, this RNA regulator appears to act post-transcriptionally (Johnsen et al. 1982; Christensen et al. 1984), whereas in Firmicutes, Cyanobacteria, and Fusobacteria, the L10$(L12)_4$-binding RNA structure is followed by an Rho-independent transcription terminator stem (Iben and Draper 2008; Yakhnin et al. 2015). Several additional ribosomal protein regulatory RNA structures specific to a subset of Firmicutes species, including Bacilli, have been computationally predicted but have yet to be experimentally validated (Yao et al. 2007).

1.2 Riboswitches

1.2.1 Anatomy of a riboswitch

Though similar to RNA *cis*-regulators of ribosomal protein synthesis, riboswitches are structured mRNA elements commonly found in the 5'-UTRs of bacterial transcripts that control gene expression upon interactions with small molecule ligands, rather than in response to proteins. Riboswitches are often more complex than their ribosomal

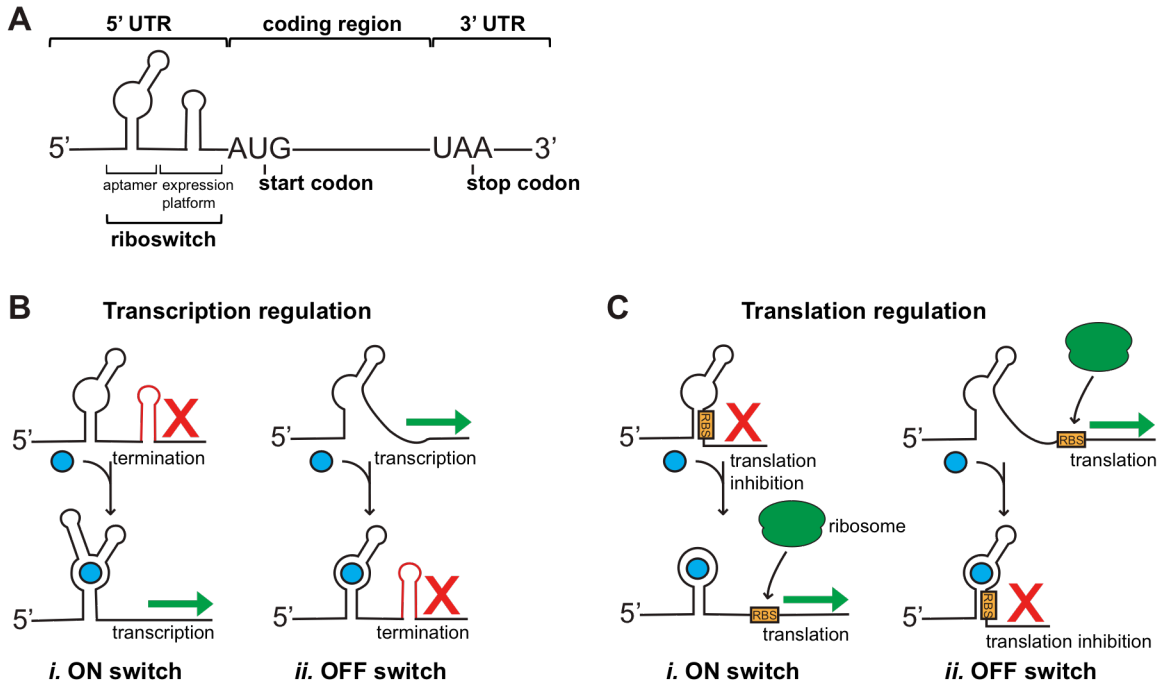


Figure 1.5. Overview of riboswitch structure and mechanisms of action. (A) A typical riboswitch is located in the 5'-UTR of an mRNA transcript and is composed of a ligand-binding domain, or aptamer, followed by an expression platform. (B) Transcriptional regulation by riboswitches. Ligand binding to the aptamer domain induces conformational changes of the expression platform that can destabilize a terminator stem to allow downstream expression ("on" switch, (i)) or stabilize a terminator to inhibit further transcription ("off" switch (ii)). (C) Translational regulation by riboswitches. Interactions between the ligand and aptamer domain result in structural reorganization of the expression platform that can expose a ribosome-binding site (RBS) to allow ribosome access and translation initiation ("on" switch, (i)) or lead to the sequestration of a RBS to prevent translation ("off" switch, (ii)).

protein-binding counterparts. They typically consist of a ligand-binding domain, or aptamer, followed by an expression platform that mediates conformational changes occurring upon ligand binding into a gene control response (Figure 1.5A). Riboswitches regulate genes important for central biochemical pathways, such as the metabolism and transport of amino acids, nucleic acids, coenzymes, and sulfur. These *cis*-regulatory RNAs commonly control expression at the transcriptional or translational level by modulating the stability of terminator/antiterminator structures (Figure 1.5B) or accessibility of ribosome-binding sites (Figure 1.5C). However, some riboswitches also utilize antisense RNAs, alternative splicing, and RNA degradation mechanisms to

regulate gene expression. The majority of riboswitches act as genetic “off” switches by inhibiting the production of biosynthetic enzymes, metabolic intermediates, or transporters when sufficient quantities are present within a cell. A subset of riboswitches function as genetic “on” switches and activate the expression of salvage or degradation pathways when their target molecules are present in excess (for review see: (Winkler and Breaker 2005; Breaker 2012; Serganov and Nudler 2013)).

1.2.2 Riboswitch diversity

Riboswitches are classified based on their target ligand as well as the conserved architectures of their aptamer domains, which can vary in length and structural complexity. Nearly 40 different classes of riboswitches that interact with a diverse array of metabolites have been characterized to date (McCown et al. 2017). Among the established riboswitch classes are three that respond to amino acids (glycine, lysine, and glutamine) (Mandal et al. 2004; Sudarsan et al. 2003; Ames and Breaker 2011), one that interacts with the amino sugar glucosamine-6-phosphate (GlcN6P) (Winkler et al. 2004), four that bind signaling molecules (cyclic di-GMP (c-di-GMP) (Sudarsan et al. 2008; Lee et al. 2010), cyclic-di-AMP (c-di-AMP) (Nelson et al. 2013), cyclic-GMP-AMP (c-GMP-AMP) (Nelson et al. 2015), and 5-aminoimidazole-4-carboxamide ribonucleoside-5’-triphosphate (ZTP) (Kim et al. 2015)), and several that recognize various enzyme cofactors, including: S-adenosylmethionine (SAM) (Grundy and Henkin 1998; Winkler et al. 2003; Breaker 2012; Gayan et al. 2018), thiamine pyrophosphate (TPP) (Winkler et al. 2002b), flavin mononucleotide (FMN) (Winkler et al. 2002a), adenosyl-cobalamin or coenzyme B_{12} (AdoCbl) (Nahvi et al. 2004), S-adenosylhomocysteine (SAH) (Wang et al. 2008; Wang and Breaker 2008), molybdenum (Moco) and tungsten (Tuco) cofactors (Regulski et al. 2008), and

tetrahydrofolate (THF) (Ames et al. 2010). Four riboswitch classes have been found to bind ions (Mg^{+2}, F^{-}, Mn^{+2}, NiCo) (Ramesh and Winkler 2017; Baker et al. 2012; Dambach et al. 2015; Furukawa et al. 2015) and several of these regulatory RNAs interact with nucleotides or nucleotide derivatives: adenine (Mandal and Breaker 2004), guanine (Batey et al. 2004), prequeuosine (preQ1) (Roth et al. 2007; Meyer et al. 2008; McCown et al. 2014), and 2'-deoxyguanosine (2'-dG) (Kim et al. 2007). In addition to small molecule metabolites, riboswitches that respond to charged tRNAs (T-boxes) to regulate amino acid and tRNA synthetase biosynthesis have also been identified (for review see: (Green et al. 2010; Zhang and Ferré-D'Amaré 2015)).

Riboswitch diversity is not limited to the wide variety of target molecules recognized. While most riboswitch sequences and structures are well conserved across different species, certain ligands can be recognized by several different aptamer conformations. The SAM riboswitches best exemplify this variation in aptamer composition; at least six distinct riboswitch structures recognize and respond to interactions with SAM and/or its derivative SAH (Wang and Breaker 2008; Breaker 2012; Gayan et al. 2018). Two different ligand-binding structures have been characterized for the c-di-GMP (Sudarsan et al. 2008; Lee et al. 2010), 2'-dG (Kim et al. 2007), and Mg^{+2} (Ramesh and Winkler 2017) riboswitch classes and three unique riboswitch aptamer conformations have been identified for both the guanidine (Nelson et al. 2017; Sherlock et al. 2017; Sherlock and Breaker 2017) and preQ1 (Roth et al. 2007; Meyer et al. 2008; McCown et al. 2014) riboswitch classes.

"Stacking" of riboswitch components adds an additional layer of complexity to riboswitch composition and allows for enhanced ligand-binding affinity and/or more dynamic genetic control. Tandem arrangements of complete and functionally independent riboswitches have been identified in select bacterial genomes. *Bacillus*

anthracis employs a double-tandem TPP riboswitch configuration (Welz and Breaker 2007) and a triple-tandem c-di-GMP riboswitch system has been characterized in a subspecies of *Bacillus thuringiensis* (Zhou et al. 2016). A tandem arrangement of two different SAM riboswitch subtypes (SAM-II/SAM-V) allows both transcriptional and translational control of downstream expression in *Candidatus pelagibacter ubique* (Poiata et al. 2009) and a conformation in which a SAM-I riboswitch precedes an AdoCbl riboswitch regulates expression in response to either ligand in *Bacillus clausii* (Nahvi et al. 2004). Moreover, the glycine riboswitch found in both *B. subtilis* and *Vibrio cholerae*, among other bacteria, consists of two homologous glycine-binding aptamers arranged in tandem followed by a single expression platform (Mandal et al. 2004); a similar architecture also has been proposed for the glutamine riboswitch (Ames and Breaker 2011). Riboswitches that bind two molecules within a single aptamer structure have also been described (Trausch et al. 2011; Gao and Serganov 2014; Ren and Patel 2014).

1.2.3 Phylogenetic distribution of riboswitches

Because riboswitch-mediated regulation exploits the bacteria-specific coupling of transcription and translation, riboswitches are almost exclusively found in bacterial genomes. The TPP riboswitch is the only known example that can be found in all three domains of life; representatives of this riboswitch class have been identified in bacteria, archaea, and eukaryotes, specifically in select plants and fungi (Barrick and Breaker 2007). The AdoCbl riboswitch is the second most prevalent class in bacteria, followed by the SAM-I riboswitch (Barrick and Breaker 2007; McCown et al. 2017). These widely distributed RNA structures and the ligands with which they interact are proposed to have ancient origins (White 1976; Benner et al. 1989; Breaker 2012). In contrast, evidence suggests the narrowly distributed riboswitches that have relatively simple structures

and/or multiple aptamer conformations, such as the preQ1 riboswitch, likely resulted from independent derivation events. Finally, the instances in which certain riboswitches occur at low frequency in select clades may be due to horizontal gene transfer. Overall, Firmicutes (Bacilli, Clostridia, Erysipelotrichia, and Negativicutes), Fusobacteria, and Gammaproteobacteria employ the largest diversity of riboswitch classes, although Gammaproteobacteria contain fewer representatives of each class (Barrick and Breaker 2007; McCown et al. 2017).

1.3 Discovery and validation of RNA *cis*-regulators

The RNA structures responsible for the control of ribosomal protein synthesis are among the earliest known examples of ligand-binding *cis*-regulatory RNA structures in bacteria. These autoregulatory RNA elements were first described in the late 1970s when overexpression of select ribosomal protein operons did not result in gene dosage effects, but rather reduced ribosomal protein production (Fallon et al. 1979; Lindahl and Zengel 1979). In the following decades, a number of additional ribosomal protein RNA *cis*-regulators were identified in *E. coli* using similar molecular genetic approaches (Zengel and Lindahl 1994). Correspondingly, the first riboswitches were discovered in the early 2000s when the expression of certain mRNAs was found to depend on the presence of specific small molecules as well as the sequence of the transcripts' 5'-UTRs (Nou and Kadner 2000; Mironov et al. 2002; Nahvi et al. 2002; Winkler et al. 2002b, 2002a).

With the advent of high-throughput sequencing technologies and the concurrent advancements made in comparative genomics and RNA structure prediction, a plethora of putative RNA *cis*-regulators have been predicted in diverse bacterial species in recent years (Barrick et al. 2004; Weinberg et al. 2007, 2010; Yao et al. 2007; Meyer et al.

2009; Xu et al. 2009; Slinger et al. 2014). A number of these RNAs have been experimentally validated, however, many remain uncharacterized. As the amount of genomic and transcriptomic data available continues to grow, it is increasingly necessary to translate computational discoveries into biologically validated RNA structures.

The majority of RNA *cis*-regulators are experimentally characterized using a combination of *in vitro* and *in vivo* techniques. Direct RNA-ligand interactions and subsequent binding kinetics are often determined *in vitro* via filter binding assays (Hall and Kranz 1999), electrophoretic mobility shift assays (Hellman and Fried 2009), inline or lead-probing (Pan 2001; Lindell et al. 2002; Regulski and Breaker 2008), and isothermal titration calorimetry (Batey et al. 2004). Structural features and binding determinants of the RNA regulator in both the presence and absence of its target molecule are further assessed using secondary structure probing methods, such as nuclease cleavage assays (Knapp 1989) and SHAPE (selective 2'-hydroxyl acylation analyzed by primer extension) (Wilkinson et al. 2006), and in more advanced instances, through X-ray crystallography (Batey et al. 2004) and nuclear magnetic resonance (Serganov et al. 2004). Regulatory activity is often examined using *in vivo* reporter systems and mechanism of action is determined through Northern blots (Rio 2015) or quantitative RT-PCR (Freeman et al. 1999), sometimes coupled with *in vitro* transcription termination (Artsimovitch and Henkin 2009) and/or translation inhibition assays (Castro-Roa and Zenkin 2015).

While the above approaches allow for the successful functional and structural characterization of putative RNA *cis*-regulators, such methods may not always capture the actual physiological behavior of an RNA structure. Discrepancies often exist between *in vitro* and *in vivo* experimental results as well as between regulatory RNA behavior within an *in vivo* reporter assay system versus within its biological context in the cell.

Limited work has examined the *in vivo* function of RNA *cis*-regulators in the context of their native loci and how these RNA elements contribute to overall cell fitness. Furthermore, despite the prevalence of RNA *cis*-regulators within bacterial genomes, little is known about how such regulatory RNA structures emerge and actively evolve, especially with their corresponding binding partners.

With that said, this body of work aims to further bridge the gap between the computational discovery and experimental verification of regulatory RNA structures to gain a more complete understanding of *cis*-regulatory RNA function and evolution in bacteria. This thesis begins with the biological validation of a widely distributed RNA *cis*-regulator of ribosomal protein synthesis that was previously identified by my group using comparative genomic approaches (Chapter 2; see Appendix for preliminary validation of additional predicted RNA regulators). In the subsequent Chapters, I move beyond the "discover-validate" binary and combine knowledge from previous *in vitro* and *in vivo* studies with an elegant knock-out strategy to assess the behavior and fitness contributions of two unique *cis*-regulatory RNAs within the context of their native loci in the model organism *B. subtilis*: the ribosomal protein L20-interacting RNA *cis*-regulator (Chapter 3) and the tandem glycine riboswitch (Chapter 4).

Chapter 2

A conserved S6:S18-interacting RNA *cis*-regulator inhibits translation of *Escherichia coli rpsF*

The content in this Chapter is adapted from the following publication:

Babina AM, Soo MW, Fu Y, Meyer MM. 2015. An S6:S18 complex inhibits translation of *E. coli rpsF*. *RNA* 21: 2039-2046.

Author contributions:

AMB performed the majority of the experiments, analyzed data, and wrote the manuscript. MWS generated select constructs and collected preliminary data. YF laid the groundwork for this project with previous *in vitro* studies. MMM conceived of the project, designed experiments, conducted the qRT-PCR analysis, analyzed data, and wrote the manuscript.

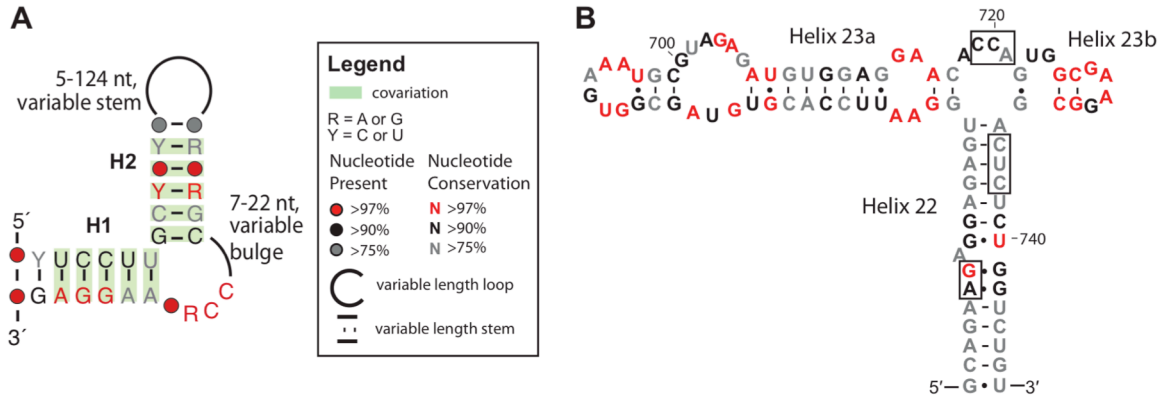


Figure 2.1. An RNA structure preceding *rpsF* is widely distributed across many bacterial phyla. (A) Consensus secondary structure of the RNA motif derived from over 1300 bacterial sequences. (B) Binding site for the S6:S18 heterodimer on 16S rRNA. Sequence is derived from *B. subtilis*; nucleotide conservation was determined from 4214 bacterial 16S rRNA sequences (Cannone et al. 2002). Black nucleotides are >90% conserved and red nucleotides are >98% conserved across bacteria. Nucleotides that make direct contact with S18 in *Thermus thermophilus* are boxed (Agalarov et al. 2000). Adapted from (Fu et al. 2014).

2.1 Introduction

We, and others, previously reported a conserved RNA structure preceding *rpsF*, which encodes ribosomal protein S6 (Matelska et al. 2013; Fu et al. 2014). The mRNA structure is widely distributed to many bacterial species and displays some similarity with the S6:S18-binding site on the 16S rRNA. In particular, both RNA structures contain a pair of highly conserved bulged cytosines flanked by two helices (Figure 2.1) (Cannone et al. 2002). In most bacteria, *rpsF* is co-localized in the genome with *priB*, which encodes a component of the primosome, and *rpsR*, which encodes ribosomal protein S18. In *E. coli*, this transcriptional unit also includes *rplI*, encoding ribosomal protein L9 (Isono and Kitakawa 1978). However, in many cases (including *B. subtilis*) *rplI* is found elsewhere in the genome. In the context of ribosome assembly, S6 and S18 form a heterodimer prior to interaction with the rRNA-S15 complex (Held et al. 1974; Agalarov et al. 2000; Recht and Williamson 2001). Examples of the conserved RNA structure preceding *rpsF* from *E. coli* and *B. subtilis* were found to specifically interact *in vitro* with

an S6:S18 heterodimer with nM affinity (Matelska et al. 2013; Fu et al. 2014), however the *B. subtilis* homolog also had weak, potentially non-specific, interactions with S18 in the absence of S6 (Fu et al. 2014). Additionally, similarities between the conserved RNA structure preceding *rpsF* and the rRNA-binding site of the S6:S18 heterodimer were identified. The combination of this *in vitro* binding data with the proximity of the RNA structure to the translation start codon strongly suggests that the conserved RNA structure allows regulation of these proteins in many species of bacteria.

In this work, we demonstrate that an example of the RNA structure from *E. coli* negatively regulates gene expression only in response to overexpression of both S6 and S18 using a *lacZ* reporter. This regulation may be disrupted by mutations to the regulatory RNA element that prevent its interaction with the repressor proteins, mutations to S18 that prevent interaction with the rRNA, and mutations to both S6 and S18 that prevent their interaction with one another. These results indicate that the S6:S18 complex is the biologically active effector. Furthermore, assessment of transcript levels by quantitative RT-PCR (qRT-PCR) suggests that the mechanism by which this RNA regulates gene expression is most likely inhibition of translation. Thus, this mRNA structure joins a collection of mRNA structures in *E. coli* that together, allow the fine-tuning of ribosomal protein levels across multiple transcriptional units.

2.2 Results and Discussion

2.2.1 rpsF_leader is a regulatory element

To assess the regulatory ability of the mRNA sequence preceding the *rpsF* gene in *E. coli* (rpsF_leader), we constructed a translational fusion between the RNA structure, including the first nine codons of *rpsF,* and *lacZ*, under transcriptional control of the IPTG-inducible Lac promoter (Maciąg et al. 2011; Slinger et al. 2014). To supply

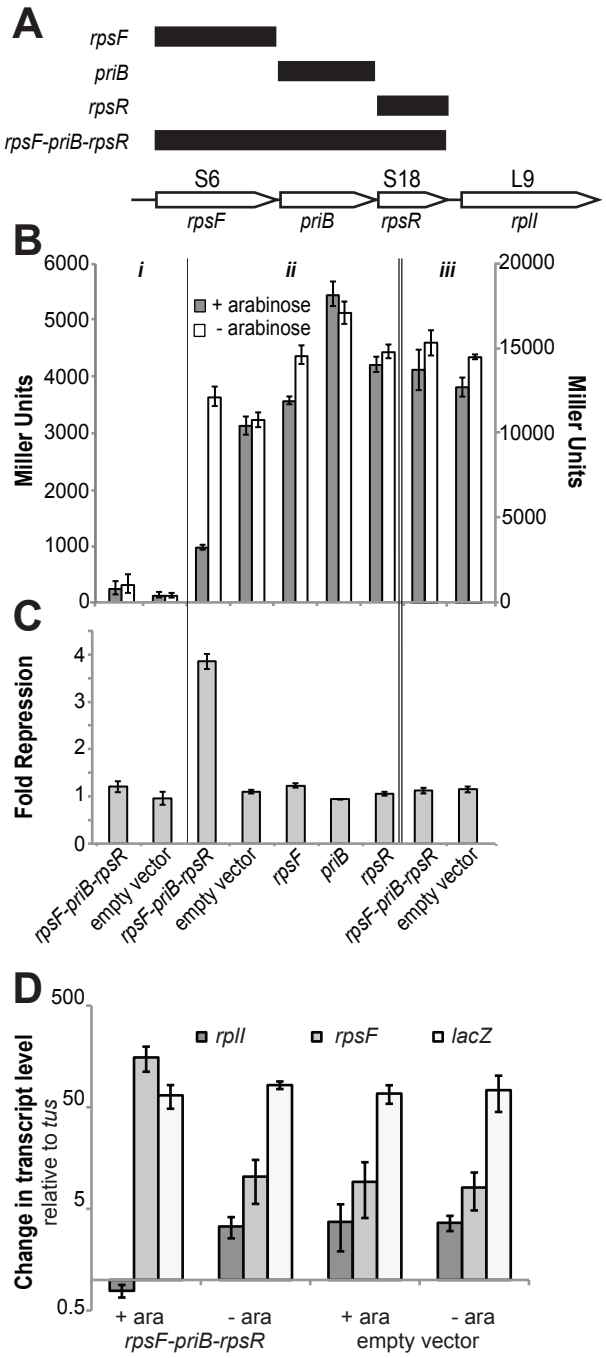


Figure 2.2. Regulatory activity of the rpsF_leader in response to *rpsF* operon overexpression. (A) Portions of the *rpsF* operon assessed in this study. (B) β-galactosidase activity (in Miller Units) of *i*: *(left axis)* cells with no *lacZ* reporter transcript induced (no IPTG); *ii: (left axis)* cells with the *rpsF'-'lacZ* transcript induced (+IPTG) and different portions of an exogenous *rpsF-priB-rpsR* transcript (including an empty vector control) induced (+arabinose) and uninduced (-arabinose); iii: *(right axis)* cells with the *rpsO'-'lacZ* transcript induced (+IPTG) with the empty vector and the *rpsF-priB-rpsR* transcript induced and uninduced (±arabinose). Values reported represent the mean of three or more independent experimental replicates. Error bars represent the standard error of the mean across experimental replicates. (C) Fold repression of the *rpsF'-'lacZ* reporter construct derived from data in B. Fold repression is calculated from matched pairs of cultures as: (Miller Units –arabinose)/(Miller Units +arabinose). Values reported represent the mean of this calculation from matched pairs of three or more independent experimental replicates. Error bars represent standard error of the mean for this calculation across experimental replicates. (D) qRT-PCR quantification of the native *rpsF-priB-rpsR-rplI* transcript (*rplI*), overexpressed transcript (*rpsF*), and reporter transcript (*lacZ*) relative to the *tus* control transcript. Values reported represent the mean of three or more independent biological replicates. Error bars represent standard error of the mean across biological replicates.

potential exogenous protein regulatory partners, we amplified portions of the *rpsF* operon and overexpressed them under the control of an arabinose-inducible promoter on pBAD33 (Figure 2.2A). These plasmids were co-transformed and the β-galactosidase activity of individual colonies was quantified with potential binding partners induced (+arabinose) and uninduced (-arabinose) (Figure 2.2B). The assays were conducted with cells harvested during

logarithmic phase growth (OD_{600} = 0.4 to 0.8) when the overexpressed protein-binding partners had been induced (+arabinose) for 2-3 hours, and the reporter construct (*lacZ*) induced for 30 minutes. β-galactosidase activity of cells where the *lacZ* transcript is uninduced is negligible (Figure 2.2Bi).

Using this system, overexpression of each individual component of the *rpsF* operon: *rpsF* encoding S6, *priB* encoding a component of the primosome, and *rpsR* encoding S18, results in little to no change in β-galactosidase activity (~1-fold repression) compared to the empty vector control (pBAD33) (Figure 2.2C). Upon expression of the first three genes of the *rpsF* operon (*rpsF-priB-rpsR*), we observe a large decrease in β-galactosidase activity (~4-fold repression). To control for potential global changes in translational efficiency that may be due to ribosome defects associated with overexpression of S6 and S18, we also examined the β-galactosidase activity of a similar *lacZ* fusion with an RNA structure not expected to interact with S6 or S18, that preceding ribosomal protein S15 in *E. coli* (*rpsO'-'lacZ*) (Philippe et al. 1993; Slinger et al. 2014). While the *rpsO'-'lacZ* fusion results in higher β-galactosidase activity in comparison to the rpsF_leader-*lacZ* constructs (Figure 2.2B), there is no significant change in β-galactosidase activity upon expression of the *rpsF-priB-rpsR* (S6:S18) construct compared to pBAD33 lacking any insert (empty vector) (Figure 2.2C). Thus, overexpression of both S6 and S18 (*rpsF-priB-rpsR*) is necessary to regulate gene expression, and the observed change in gene expression is specific to the rpsF_leader.

2.2.2 rpsF_leader inhibits translation

The *E. coli* example of the rpsF_leader contains a putative Shine-Dalgarno (SD) sequence within the structure, suggesting that the mechanism of action for this RNA is through inhibition of translation initiation. To examine the mechanism of regulation, we

measured mRNA levels in our reporter strains via qRT-PCR to determine whether transcript levels are significantly altered under the conditions where we observe changes in reporter gene expression. The *tus* gene (terminus utilization substance) was used as a control transcript instead of a ribosomal protein or rRNA control transcript, as they might be affected by the overexpression of the *rpsF-priB-rpsR* fragment (Sykes et al. 2010). We measured levels of the overexpressed transcript (primers within *rpsF*), the reporter transcript (primers within *lacZ*), and the native transcript (primers within *rplI*) relative to *tus* transcript. We found that levels of the *lacZ* transcript do not change relative to the *tus* transcript under the +arabinose condition where we observe changes in β-galactosidase activity (Figure 2.2D)**.** In contrast, while it is clear that the overexpression constructs increase *rpsF* transcript levels approximately 10-fold upon induction with arabinose as expected, the *rplI* transcript (corresponding to the native *rpsF-priB-rpsR-rplI* transcript) is reduced approximately 4.5-fold under these conditions. These results indicate that while the changes in β-galactosidase activity we observe are due to differences in translation, the native transcript is subject to additional regulation and is either not produced, terminates prematurely, or is degraded more rapidly under these conditions.

2.2.3 Mutations to rpsF_leader affect expression and regulatory capacity

To determine whether mutations to the RNA that prevent protein binding could abolish regulation, we examined six mutant RNAs (Figure 2.3A). The M1 mutation disrupts the H1 stem as well as the putative SD sequence. This mutation in the homologous *B. subtilis* RNA was previously shown to reduce protein binding affinity using *in vitro* electrophoretic mobility shift assays (EMSAs) (Fu et al. 2014). During our *in vivo* studies, we find that this mutation results in low β-galactosidase activity in both the

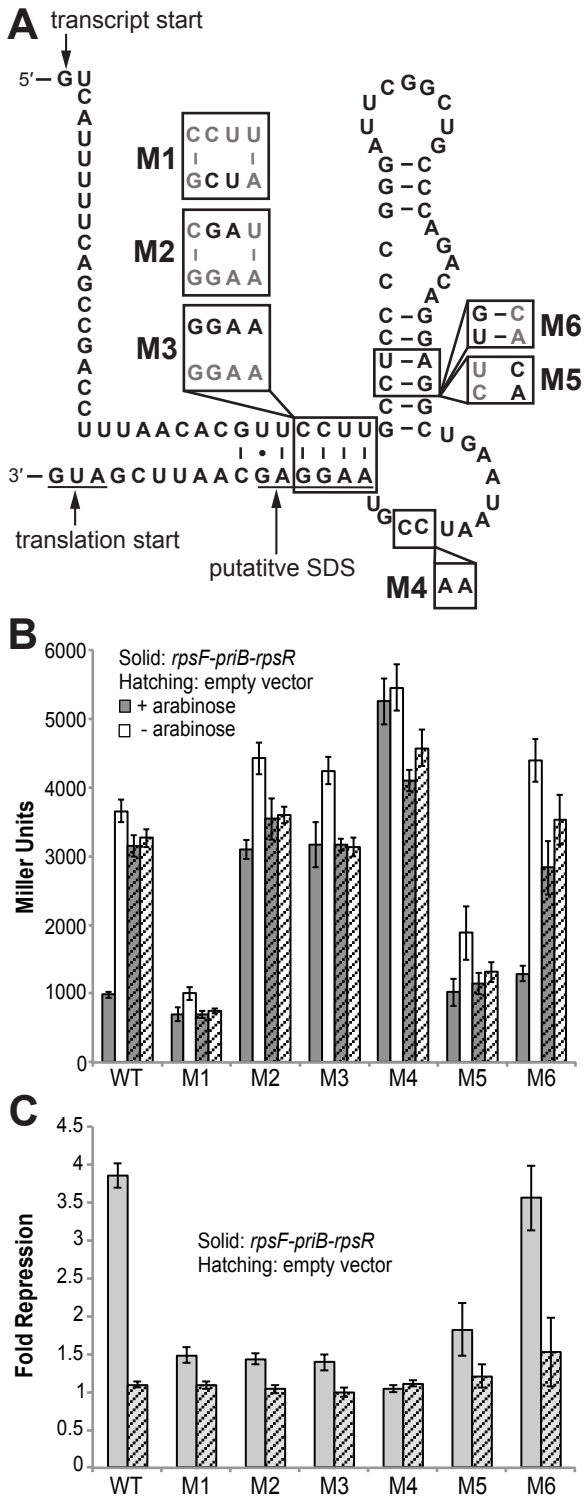


Figure 2.3. Regulatory capacity of rpsF_leader mutants. (A) The presumed secondary structure of the *rpsF* 5'-UTR used for reporter studies with mutations M1-M6. The transcription start site (Maciag et al. 2001), translational start, and putative Shine-Dalgarno (SD) sequence are indicated. (B) β-galactosidase activity (in Miller Units) of cells carrying plasmids with the unmutated rpsF_leader (WT) or each mutant RNA (M1-M6) and the *rpsF-priB-rpsR* overexpression plasmid or the empty vector (pBAD33) under induced and uninduced (±arabinose) conditions. Values reported represent the mean of three or more independent experimental replicates. Error bars represent the standard error of the mean across experimental replicates. (C) Fold repression as calculated in Figure 2.2C of the unmutated rpsF_leader (WT) and each mutant RNA (M1-M6). Values reported represent the mean of this calculation from matched pairs of three or more independent experimental replicates. Error bars represent standard error of the mean for this calculation across experimental replicates.

presence and absence of exogenous protein (Figure 2.3B). This indicates that this region is important for translational efficiency and supports our prediction that the H1 stem may contain the SD sequence. The overall low β-galactosidase activity of M1 makes any potential regulatory activity difficult or impossible to determine.

The M2 mutation is not directly in the proposed protein-binding region but is within positions that are predicted to pair with the SD sequence. This mutation in the homologous *B. subtilis* RNA completely abolished

protein binding *in vitro* (Fu et al. 2014). Consistent with these data, we find that this mutation to the *E. coli* rpsF_leader almost completely abolishes regulation. Further disruption of this stem with the M3 mutation yields similar results, suggesting that the base pairing in this region is important for protein binding and consequent regulation (Figure 2.3C).

The M4 mutation is directly within the proposed protein-binding site, changing two highly conserved cytosines to adenines (Matelska et al. 2013; Fu et al. 2014). During *in vitro* studies of the homologous *B. subtilis* RNA, this mutation significantly inhibited protein binding. A more severe mutation at the same position (AAA rather than AAG) in the *E. coli* RNA had a similar effect during *in vitro* binding assays (Matelska et al. 2013). We find that this mutation completely abolishes the demonstrated regulation *in vivo* (Figure 2.3C). This is consistent with its location directly within the protein-binding region.

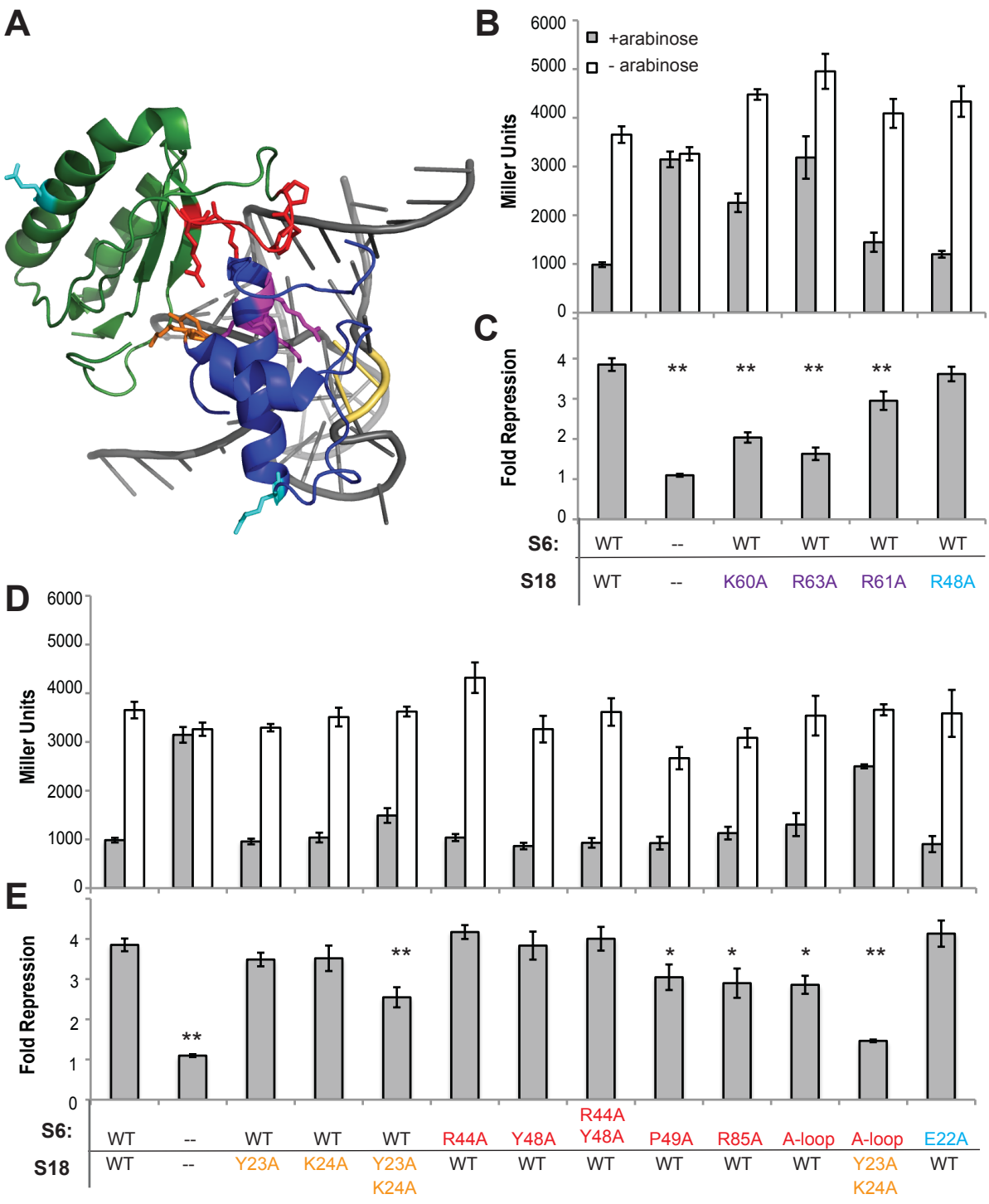
The M5 mutation disrupts the highly conserved H2 stem of the rpsF_leader secondary structure. Mutations to this region in the homologous *B. subtilis* RNA significantly impaired RNA-protein interactions *in vitro* (Fu et al. 2014). In agreement with these data, this mutation reduces repression of β-galactosidase activity to about half of that observed with the wild-type rpsF_leader (~1.8-fold repression). The low basal (unrepressed) β-galactosidase activity of the M5 mutant in comparison to the wild-type rpsF_leader and other mutant constructs (Figure 2.3B) does not appear to be due to changes in transcript levels (data not shown), but rather to translational efficiency. The M6 compensatory mutation to the H2 stem almost fully restores basal β-galactosidase activity and regulation to near wild-type levels (~3.5-fold repression) (Figure 2.3B,C). This further confirms our previous secondary structure predictions and indicates that the base pairing in the H2 stem is important for regulatory activity.

2.2.4 S6 and S18 residues required for regulation

The S6:S18 complex is expected to be the biologically active regulator that interacts with the rpsF_leader. During ribosome assembly, S6 and S18 form a heterodimer prior to assembly with the 16S rRNA-S15 complex (Held et al. 1974; Recht and Williamson 2001). The rpsF_leader shows significant sequence and structural similarities with the S18-binding site on the rRNA (Figure 2.1), and tertiary structure modeling indicates that the majority of RNA-protein contacts are likely with S18 (Matelska et al. 2013). *In vitro* studies with the *B. subtilis* rpsF_leader homolog show weak and likely non-specific interactions between S18 and the rpsF_leader (Fu et al. 2014), and our β-galactosidase assay results (Figure 2.2B,C) indicate that overexpression of both S6 and S18 is required for inhibition.

To assess the role S18 has in the mRNA-protein regulatory interactions, we altered two positively charged amino acids in S18 (within the *rpsF-priB-rpsR* construct) that are expected to interact with the conserved pair of cytosines in both the rRNA and

Figure 2.4. rpsF_leader regulation in response to S6:S18 protein mutations. (A) Rendering of the S6:S18 heterodimer in complex with the 16S rRNA (coordinates derived from 2QAL (Borovinskaya et al. 2007)). The rRNA segment (660-678; 713-739) is gray, interacting bases C719 and C720 are highlighted in yellow, S18 is shown in blue, amino acids mutated at the S18:rRNA interface (K60, R61, and R63) are indicated in purple, amino acids mutated at the S6:S18 interface (Y23 and K24) are orange. S6 is displayed in green, and the amino acids mutated at the S6:S18 interface are highlighted in red. For individual amino acids mutated at the S6:S18 interface (R44, Y48, P49 and R85) side chains are displayed, for the additional amino acids altered in the 'A-loop' mutant (44-49 all mutated to alanine) only the backbone is colored. Negative control mutations (S6 E22 and S18 R48) are highlighted in cyan. (B) β-galactosidase activity (in Miller Units) of cells carrying the *rpsF-priB-rpsR* overexpression construct with mutations to the S18 RNA-binding region with protein induced and uninduced (±arabinose). For comparison, data for the unmutated construct (WT) and empty vector (--) are included. Values reported represent the mean of three or more independent experimental replicates. Error bars represent the standard error of the mean across experimental replicates. (C) Fold repression for S18 RNA-binding site mutations (data in B) calculated as described in Figure 2.2C. ** indicates a statistically significant change ($p<0.01$) from the wild-type construct. (D) β-galactosidase activity (in Miller Units) of cells carrying the *rpsF-priB-rpsR* overexpression construct with mutations made to the S6:S18 interface. For comparison, data for the unmutated construct (WT) and empty vector (--) are included. Values reported represent the mean of three or more independent experimental replicates. Error bars represent standard error of the mean across experimental replicates. (E) Fold repression for S6:S18 interface mutations (data in D) calculated as described in Figure 2.2C. Asterisks indicate a statistically significant change from the wild-type construct: ** indicates $p<0.01$ and * indicates $p<0.05$.


A
B
+arabinose
- arabinose
Miller Units
6000
5000
4000
3000
2000
1000
0
C
Fold Repression
4
3
2
1
0
**
S6:
S18
WT
--
WT
WT
WT
WT
WT
--
K60A
R63A
R61A
R48A
D
Miller Units
E
Fold Repression
**
*
S6:
S18
WT
--
WT
WT
WT
R44A
Y48A
R44A
Y48A
P49A
R85A
A-loop
A-loop
E22A
WT
--
Y23A
K24A
Y23A
K24A
WT
WT
WT
WT
WT
WT
Y23A
K24A
WT

rpsF_leader (lysine 60 and arginine 63 to alanine) (Figure 2.4A). These amino acids form hydrogen bonds with the conserved cytosines in the loop region adjacent to Helix 23a of the rRNA (C719 and C720) (Figure 2.1B) (Agalarov et al. 2000) that are mimicked in the mRNA structure by the conserved pair of cytosines mutated in M4 (Figure 2.3A). During *in vitro* studies, mutating these amino acids resulted in either significant reduction in binding affinity (K60A) or complete loss of saturated binding (R63A) (Matelska et al. 2013). In our assays, both of these mutants significantly reduce regulation (~2-fold repression, $p<0.01$) (Figure 2.4B,C). Additionally, mutating arginine 61 to an alanine, a residue that is implicated in both S18-rRNA interactions and S6:S18 dimer interactions in a *Thermus thermophilus* structure of the S15:S6:S18:rRNA complex (Agalarov et al. 2000), but only appears to contact the RNA in crystal structures of the *E. coli* ribosome (Borovinskaya et al. 2007), decreases regulatory activity (~2.9-fold repression, $p<0.01$). This supports the conclusion that S18 contacts with the mRNA are critical for binding and subsequent regulatory activity. Arginine 48 was also mutated to an alanine as a negative control. R48 is not known to directly contribute to S18-rRNA or S6:S18 protein-protein interactions, and this mutation to S18 did not affect regulation.

To further assess whether the S6:S18 heterodimer is the biologically relevant effector, we mutated residues in both S6 and S18 (within the *rpsF-priB-rpsR* construct) to disrupt S6:S18 dimer interactions and consequent regulation. However, the interface of S6 and S18 proved difficult to completely disrupt using mutations to a single protein. Mutations to S18, tyrosine 23 and lysine 24 to alanine, individually had little or no effect on regulation. However, combining these mutations results in a modest, but significant ($p<0.01$) decrease in β-galactosidase activity (2.5-fold repression). On the complementary S6 surface, individual and combined mutations of S6 arginine 44 and tyrosine 48 to alanine had little effect on binding (Figure 2.4D,E). Mutations of proline 49

and arginine 85 to alanine, and replacing residues 44-49 (RQLAYP) with alanine ('A-loop') all had modest, but significant effects on repression (decreasing to about ~3-fold, $p<0.05$). However, combining the S6 'A-loop' mutation with the S18 Y23A/K24A mutation did strongly impact regulation (1.5-fold repression, $p<<0.001$), indicating that disruption of the S6:S18 interface can disrupt regulation. A negative control mutation to S6, glutamate 22 to alanine, showed no change in repression.

We suspect the robustness of the protein-protein interaction is due to several factors. First, our overexpressed mutant proteins are competing with endogenous protein levels, and we may see less of an effect due to this competition. Second, the nature of the S6:S18 protein-protein interaction may be somewhat plastic and robust to our mutagenesis efforts. The structures of the *T. thermophilus* S15:S6:S18:rRNA complex and the complete *E. coli* ribosome are not identical, leading to slightly varying determinations of which residues are most critical for this interaction (Agalarov et al. 2000; Borovinskaya et al. 2007). In particular the N-terminal portion of S18 that contacts S6 is not resolved in the *T. thermophilus* complex, suggesting that there may be some flexibility in the S6:S18 interaction. That our combined S6 'A-loop' and S18 Y23A/K24A mutations had the strongest impact on regulation indicates that the first of these two factors is likely playing a significant role, and supports our conclusion that the active complex is the S6:S18 heterodimer.

2.3 Conclusions

In this work we show that in *E. coli,* the rpsF_leader RNA is a regulatory element that inhibits the translation of *rpsF*. Changes in β-galactosidase activity are not accompanied by corresponding changes in the level of *lacZ* transcript. However, the native transcript does show significant reduction under conditions where β-galactosidase

activity is reduced, suggesting that it is subject to additional regulation. Due to past associations of reduced translation with more rapid transcript degradation (Deana and Belasco 2005), we strongly suspect this mechanism may play a significant role in amplifying the relatively modest 4-fold repression we observe.

Our mutagenesis studies indicate that the biologically relevant effector is the S6:S18 heterodimer. Unlike many regulators of ribosomal protein synthesis, the rpsF_leader does not respond to a primary rRNA-binding protein, but rather a complex of secondary rRNA-binding proteins. This situation is relatively rare, as most characterized regulators to date interact with primary rRNA-binding proteins, although there are a few exceptions (e.g. S2, L25 (Aseev et al. 2008, 2015)). The closest comparison to this situation is the *rplJ-rplL* regulator that interacts with either L10 or the L10(L12)$_4$ complex (Yates et al. 1981).

The rpsF_leader is one of the few RNA structures responsible for ribosomal protein regulation that is widely distributed across many bacterial phyla (Matelska et al. 2013; Fu et al. 2014). For other such widely distributed regulators, the same protein-binding site may be utilized for different mechanisms of gene regulation. For example as mentioned in Chapter 1, the L10(L12)$_4$-interacting regulatory mRNA inhibits translation initiation in *E. coli* (Johnsen et al. 1982; Christensen et al. 1984), whereas in *B. subtilis* the conserved L10(L12)$_4$ mRNA-binding site regulates transcription attenuation (Iben and Draper 2008; Yakhnin et al. 2015). This is similar to many riboswitch aptamers where the same recognizable aptamer may utilize different expression platforms in different species (Barrick and Breaker 2007). Therefore, it would not be unlikely for the rpsF_leader to utilize distinct mechanisms for regulation in diverse species of bacteria. Confirmation of the biological activity of the rpsF_leader mRNA structure allows it to join a still growing canon of RNA regulatory structures in bacteria that allow regulation of

ribosomal protein synthesis, but utilize a wide variety of mechanisms and have very different distribution patterns across bacterial phyla.

2.4 Materials and Methods

2.4.1 Plasmid construction

The *lacZ* reporter plasmid is a modified version of the pLac-ThiM#2-tetA-gfpuv plasmid previously described (Muranaka et al. 2009; Slinger et al. 2014), in which the tetA-gfpuv construct is replaced with a *lacZ* gene using SalI and XbaI restriction sites. The fragment corresponding to the RNA and the first 9 codons of *rpsF* (NC_000913.3: 4425023-4425144) was PCR-amplified from genomic DNA extracted from *E. coli* strain XL1-Blue (Agilent Technologies) and cloned between the EcoRI and SalI sites, to generate an RNA-*lacZ* translational fusion (Table 2.1). To construct the pBAD33 (ATCC 87402) overexpression plasmids, genomic fragments corresponding to portions of the *rpsF* operon (NC_000913.3 *rpsF*: 4425118-4425513; *priB*: 4425520-4425834; *rpsR*: 4425839-4426066; *rpsF-priB-rpsR*: 4425118-4426066) were PCR-amplified from genomic DNA from *E. coli* and cloned between the XbaI and SacI sites. To enable efficient translation, a Shine-Dalgarno sequence (5'-AGGAGGTTTTAAA) was appended to the 5' end of each genomic fragment. Mutant RNA and protein plasmids were created by site-directed mutagenesis from the original plasmids (Table 2.1). All plasmid sequences were confirmed by Sanger sequencing.

2.4.2 β-galactosidase activity assays

To generate strains for β-galactosidase activity analysis, a reporter plasmid and a protein overexpression plasmid were co-transformed into *E. coli* strain NCM534 (F-, *Δ(araD-araB)714*, *Δ(lacA-lacZ)880(::FRT)*, *lacIp-4000(lacI^{Q})*, *zah-2225::FRT*, *λ^{-}*, *ΔaraEp-*

532::FRT, *φP$_{cp18}$araE533*, *rph-1*) (*E. coli* Genetic Stock Center, strain #: 8256). For each independent assay, a single colony was chosen and grown overnight at 37ºC with shaking (225 rpm) in 1.5 mL LB + ampicillin (100 μg/mL) + chloramphenicol (35 μg/mL). This culture was used to inoculate two separate pre-warmed 1.5 mL cultures, one containing 15 mM L-arabinose (+arabinose) to induce protein overexpression and one without any sugar added (-arabinose, protein not induced), to a starting OD_{600} of approximately 0.05-0.1. Cultures were grown at 37ºC with shaking (225 rpm) until an OD_{600} of approximately 0.4-0.6 was reached. IPTG (1 mM) was added to both cultures to induce transcription of the reporter transcript. The induced cultures were incubated for 30 minutes at 37ºC with shaking (225 rpm). Spectinomycin (100 μg/mL) was added to each culture following the 30-minute incubation to inhibit further translation. Cells (1 mL) were harvested and re-suspended in 1 mL Z buffer (50 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$, 50 mM 2-Mercaptoethanol) + 100 μg/mL spectinomycin. OD_{600} readings were recorded as cell suspensions in Z buffer. Samples that had an OD_{600} reading of 0.4 or lower at time of harvest were discarded. β-galactosidase activity assays were performed as previously described using 30 μl of cell suspensions and Miller Units were calculated as follows (Miller 1992):

Miller Units = $1000 * \frac{A_{420}}{\Delta t\ (\text{min.}) * A_{600} * vol.(mL)}$

To determine the fold repression for each sample, the Miller Units of the -arabinose (protein not induced) culture were divided by that of the corresponding +arabinose (protein induced) culture. The values reported represent 5 or more independent replicates for wild type, mutant rpsF_leader, and protein overexpression co-transformed strains, and 3 independent replicates for the uninduced (-IPTG) and *rpsO'-'lacZ* control assays. Error bars represent standard error of the mean across biological replicates. To

determine significance, the fold repression of samples was compared using a Welch's T-test in Microsoft Excel.

2.4.3 Quantitative RT-PCR

Total RNA was extracted using Trizol (Life Technologies) from +/- arabinose cultures grown essentially as described. Contaminating DNA was removed from 10 μg of total RNA by RQ1 DNase digestion at 37ºC for 4 hours (Promega) followed by phenol-chloroform extraction and ethanol precipitation. Reverse transcription reactions were conducted with random hexamers using SuperScript III (Life Technologies) on 2.5 μg of DNase-treated RNA according to the manufacturer's instructions. The resulting cDNA was used as template for qPCR using an Applied Biosystems 7500 Fast Real-Time PCR system (SYBR Green detection, ThermoFisher Scientific). qPCR primers targeting the *lacZ* reporter, the overexpression construct (*rpsF*), and the native transcript (*rplI*) were used to quantify transcript levels in each sample. Quantification of *tus* was used as a normalization control (Table 2.1). Equivalent experiments were conducted on reactions lacking reverse transcriptase to ensure that DNase digestion removed all contaminating DNA. Each condition is represented by three independent biological replicates, and qPCR was conducted with three technical replicates for each biological replicate. Standard error reported represents variance among biological replicates.

Table 2.1. Oligonucleotides used in this study.
For primer pairs, the forward primer is listed first and the reverse primer is listed second.

Name	Sequence (5'-3')	Notes
627	CGCG**GAATTC**GCGTCATTTTTCAGCCG ACCTTTAAC	Primer for cloning rpsF_leader into modified pLac reporter vector (Slinger et al. 2014); EcoRI restriction site in bold
628	ACGC**GTCGAC**CATAAAAACGATTTCGTA ATGACGCAT	Primer for cloning rpsF_leader into modified pLac reporter vector (Slinger et al. 2014); SalI restriction site in bold
629	CAA**GAGCTC**AGGAGGTTTTAAAATGCGT CATTACGAAATCGTTTTTATG	Primer for cloning *rpsF* into pBAD33 protein overexpression vector; SacI restriction site in bold; SD sequence is underlined
630	CAA**TCTAGA**TTACTCTTCAGAATCCCCA GCTTCAGC	Primer for cloning *rpsF* into pBAD33 protein overexpression vector; XbaI restriction site in bold
782	CAA**GAGCTC**AGGAGGTTTTAAAATGACC AACCGTCTGGTG	Primer for cloning *priB* into pBAD33 protein overexpression vector; SacI restriction site in bold; SD sequence is underlined
783	CAA**TCTAGA**CTAGTCTCCAGAATCTATC AATTCAAT	Primer for cloning *priB* into pBAD33 protein overexpression vector; XbaI restriction site in bold
631	CAA**GAGCTC**AGGAGGTTTTAAAATGGC ACGTTATTTCCGTCGTCGC	Primer for cloning *rpsR* into pBAD33 protein overexpression vector; SacI restriction site in bold; SD sequence is underlined
632	CAA**TCTAGA**TTACTGATGGCGATCAGTG TACGG	Primer for cloning *rpsR* into pBAD33 protein overexpression vector; XbaI restriction site in bold
667	CTGAATAATCCGTATCGAGCAATTCG	Mutagenesis primer for rpsF_leader M1
668	CGAATTGCTCGATACGGATTATTCAG	Mutagenesis primer for rpsF_leader M1
669	CCTTTAACACGTTCGATGCCTCCCCGG G	Mutagenesis primer for rpsF_leader M2
670	CCCGGGGAGGCATCGAACGTGTTAAAG G	Mutagenesis primer for rpsF_leader M2
1135	GACCTTTAACACGTTGGAAGCCTCCCC GGGATTC	Mutagenesis primer for rpsF_leader M3

1136	GAATCCCGGGGAGGCTTCCAACGTGTTAAAGGTC	Mutagenesis primer for rpsF_leader M3
671	GGAGGCTGAATAATAAGTAAGGAGCAATTCG	Mutagenesis primer for rpsF_leader M4
672	CGAATTGCTCCTTACTTATTATTCAGCCTCC	Mutagenesis primer for rpsF_leader M4
1137	GCTGACCCAGACAGGCAGCTGAATAATCCGTAAG	Mutagenesis primer for rpsF_leader M5
1138	CTTACGGATTATTCAGCTGCCTGTCTGGGTCAGC	Mutagenesis primer for rpsF_leader M5
1139	TAACACGTTCCTTGCTGCCCCGGGATTCG	Mutagenesis primer for rpsF_leader M6; use M5 as template
1140	CGAATCCCGGGGCAGCAAGGAACGTGTTA	Mutagenesis primer for rpsF_leader M6; use M5 as template
710	CTGGCTCGCGCTATCGCGCGCGCTCGCTACCTG	Mutagenesis primer for S18 K60A
711	CAGGTAGCGAGCGCGCGCGATAGCGCGAGCCAG	Mutagenesis primer for S18 K60A
708	CTATCAAACGCGCTGCGTACCTGTCCCTGCTG	Mutagenesis primer for S18 R63A
709	CAGCAGGGACAGGTACGCAGCGCGTTTGATAG	Mutagenesis primer for S18 R63A
1133	GCTCGCGCTATCAAAGCCGCTCGCTACCTGTC	Mutagenesis primer for S18 R61A
1134	GACAGGTAGCGAGCGGCTTTGATAGCGCGAGC	Mutagenesis primer for S18 R61A
1131	GCCGTATCACCGGTACCGCTGCAAAATACCAGCGT	Mutagenesis primer for S18 R48A
1132	ACGCTGGTATTTTGCAGCGGTACCGGTGATACGGC	Mutagenesis primer for S18 R48A
852	GACTGGGGCCGCGCTCAGCTGGCTTACC	Mutagenesis primer for S6 R44A
853	GGTAAGCCAGCTGAGCGCGGCCCCAGTC	Mutagenesis primer for S6 R44A
1213	CAGCTGGCTGCCCCGATCAAC	Mutagenesis primer for S6 Y48A
1214	GTTGATCGGGGCAGCCAGCTG	Mutagenesis primer for S6 Y48A
1245	CGTCAGCTGGCTTACGCGATCAACAAACTG	Mutagenesis primer for S6 P49A
1246	CAGTTTGTTGATCGCGTAAGCCAGCTGACG	Mutagenesis primer for S6 P49A
1243	GATGCCGTTATCGCCAGCATGGTTATG	Mutagenesis primer for S6 R85A
1244	CATAACCATGCTGGCGATAACGGCATC	Mutagenesis primer for S6 R85A
1241	CGCGCTGCGGCGGCTGCCGCGATCAACAAACTGCAC	Mutagenesis primer for S6 ‘A-loop’
1242	GATCGCGGCAGCCGCCGCAGCGCGGCCCCAGTCTTC	Mutagenesis primer for S6 ‘A-loop’

1141	GTTCCGGGCATGATCGCGCGCTACACT GCTGC	Mutagenesis primer for S6 E22A
1142	GCAGCAGTGTAGCGCGCGATCATGCCC GGAAC	Mutagenesis primer for S6 E22A
1235	CAAGAGATCGACGCTAAAGATATCGC	Mutagenesis primer for S18 Y23A
1236	AGCGATATCTTTAGCGTCGATCTCTTG	Mutagenesis primer for S18 Y23A
1237	GAGATCGACTATGCAGATATCGCTACG	Mutagenesis primer for S18 K24A
1238	CGTAGCGATATCTGCATAGTCGATCTC	Mutagenesis primer for S18 K24A
1239	GAGATCGACGCTGCAGATATCGCTACG CTGAAAAAC	Mutagenesis primer for S18 Y23A K24A
1240	AGCGATATCTGCAGCGTCGATCTCTTGA ACGCC	Mutagenesis primer for S18 Y23A K24A
1204	TACCTGTTCCGTCATAGCGA	Primer for qRT-PCR targeting the *lacZ* reporter
1205	CTGTTTACCTTGTGGAGCGA	Primer for qRT-PCR targeting the *lacZ* reporter
1218	GGCTTACCCGATCAACAAAC	Primer for qRT-PCR targeting the *rpsF* overexpression construct
1219	CGGAAGGTAGTTTCCAGCTC	Primer for qRT-PCR targeting the *rpsF* overexpression construct
1127	TACCATCGCGTCTAAAGCTG	Primer for qRT-PCR targeting the *rplI* native transcript
1128	TTCGCTCTTAGCCACTTCAA	Primer for qRT-PCR targeting the *rplI* native transcript
981	TGTTTTCGAAGCGACAGATG	Primer for qRT-PCR targeting the *tus* normalization control
982	TTTCGAGGCCGAGAATTTTA	Primer for qRT-PCR targeting the *tus* normalization control

Chapter 3

Fitness advantages conferred by the L20-interacting RNA *cis*-regulator of ribosomal protein synthesis in *Bacillus subtilis*

The content in this Chapter is adapted from the following unpublished manuscript:

Babina AM, Parker DJ, Li G-W, and Meyer MM. Fitness advantages conferred by the L20-interacting RNA *cis*-regulator of ribosomal protein synthesis in *Bacillus subtilis*.

Author contributions:

AMB designed and performed the experiments, analyzed data, and wrote the manuscript. DJP and G-WL assisted with the ribosome sedimentation profiling and analyzed data. MMM conceived of the project, designed experiments, analyzed data, and wrote the manuscript.

3.1 Introduction

Since the late 1970s, a number of ribosomal protein RNA *cis*-regulators have been identified and experimentally validated in bacteria and, as detailed in Chapters 1 and 2, the methods for characterizing these RNA-protein regulatory interactions are well established. These studies and techniques have greatly improved our knowledge of the structures and mechanisms responsible for the RNA-based autoregulation of ribosomal protein synthesis in bacteria. However, our understanding of the selective pressures that drive the emergence and maintenance of ribosomal protein RNA *cis*-regulators within bacterial genomes remains limited. The presence of both highly conserved and independently derived *cis*-regulatory RNA structures suggests there exists significant evolutionary pressure to regulate ribosomal protein synthesis in this manner. Additionally, although it is well known that overexpressing select ribosomal proteins inhibits bacterial growth, these observations are typically from studies that induced protein overexpression using multiple gene copies, plasmids, and/or strong promoters (Dean and Nomura 1980; Sykes et al. 2010). Little to no work has investigated the impacts of simply removing ribosomal protein autoregulation in bacteria.

To measure the fitness advantages ribosomal protein regulatory RNA structures confer to an organism, we introduced point mutations into the native locus of the L20-interacting RNA *cis*-regulator in the *B. subtilis* genome to disrupt ribosomal protein binding and regulation, and subsequently assayed the strains for mutant phenotypes. *B. subtilis* is a good model organism for examining RNA *cis*-regulators within the context of their native loci. This bacterium utilizes a variety of *cis*-regulatory RNA structures to control gene expression (Barrick and Breaker 2007; Fu et al. 2013; Deiorio-Haggar et al. 2013; McCown et al. 2017), it is naturally competent during late log phase growth (Dubnau 1991), and it is highly amenable to manipulation of its genome via homologous

recombination (Fernández et al. 2000). Furthermore, the L20-interacting regulatory RNA found in *B. subtilis* is a good subject for this study as its structure and mechanism of action are well characterized. As mentioned in Chapter 1, this RNA structure is narrowly distributed to the Firmicutes phylum and it regulates expression of the *infC-rpmI-rplT* operon (encoding translation initiation factor IF3 and large subunit ribosomal proteins L35 and L20, respectively) at the transcriptional level using a Rho-independent terminator that is stabilized upon L20 binding (Choonee et al. 2007; Bruscella et al. 2011; Deiorio-Haggar et al. 2013).

We find that removing RNA-mediated autoregulation results in elevated transcript levels of downstream genes and causes cold-sensitive defects in growth, rRNA processing, and ribosome subunit sedimentation. Our results suggest that improper regulation of ribosomal protein expression compromises ribosome biosynthesis and demonstrate the significant role *cis*-regulatory RNA structures have in proper ribosome production and overall organismal fitness. This work gives insight into why RNA-based regulation of ribosomal proteins is so prevalent across diverse bacterial species and sheds light on the selective forces that govern structured RNA evolution and conservation.

3.2 Results

3.2.1 Reporter assays confirm behavior of L20-interacting RNA mutations

We first verified the impact of specific mutations on the regulatory activity of the L20-interacting RNA and downstream gene expression using β-galactosidase reporter assays in *B. subtilis* (Figure 3.1A). Assays were conducted during log phase growth while overexpressing the entire *infC* operon from a plasmid, or with an empty plasmid. An ~11-fold reduction in β-galactosidase activity was observed for the wild-type L20-

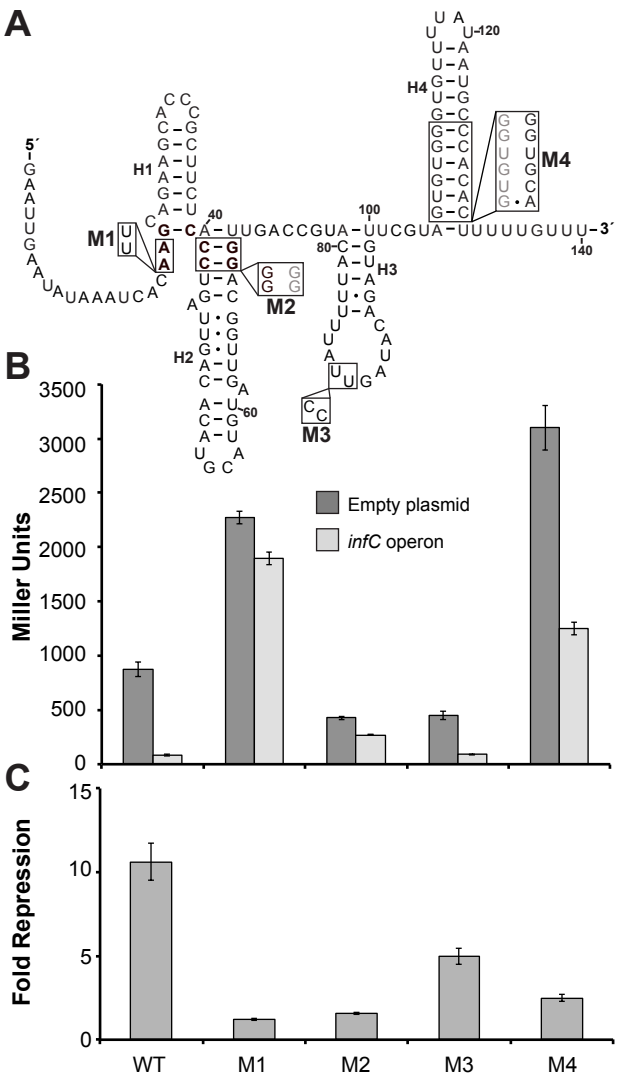


Figure 3.1. Regulatory activity of the *B. subtilis* L20-interacting RNA mutants examined in this study. (A) Secondary structure of the L20-interacting RNA in its protein-bound form with mutations M1-M4. The L20-binding site is in bold. Helix H4 is the Rho-independent transcription terminator that forms upon L20 binding. Nucleotides are numbered from the transcript start site from the second *infC* operon promoter, +1 (Choonee et al. 2007). (B) β-galactosidase activity (in Miller Units) of the L20-interacting RNA mutant constructs with overexpression of the *infC* operon or empty plasmid during log phase growth at 37ºC. The values reported represent the mean of three or more independent experimental replicates; error bars represent standard error of the mean across biological replicates. (C) Fold repression of the L20-interacting RNA reporter constructs derived from the data in B. Fold repression was calculated for each reporter construct as follows: (Mean Miller Units of empty plasmid strain)/(Mean Miller Units of *infC* operon overexpression strain). Error bars represent standard error of the mean propagated from the values in B using standard calculations (Taylor 1997).

interacting RNA construct when the *infC* operon was overexpressed (Figure 3.1B,C). This fold repression is comparable to what has been previously measured for the L20-interacting RNA (Choonee et al. 2007; Bruscella et al. 2011).

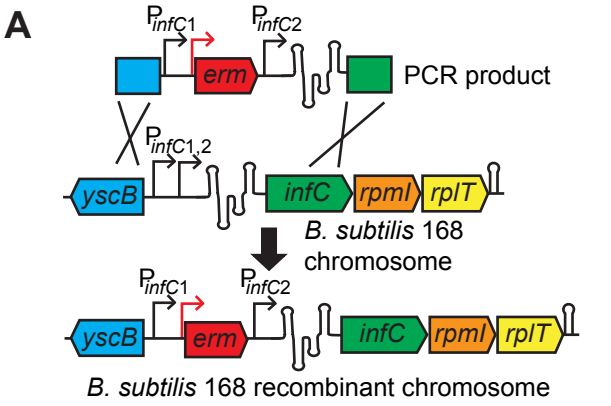
Mutations M1 and M2 disrupt the previously identified and highly conserved L20-binding site at the junction of Helices H1 and H2 (Figure 3.1A) (Choonee et al. 2007; Deiorio-Haggar et al. 2013). Both mutations relieve the repression observed with the wild-type RNA construct when the *infC* operon is overexpressed (Figure 3.1B,C). Mutation M3 targets the loop region of Helix H3 and serves as a control. Previous

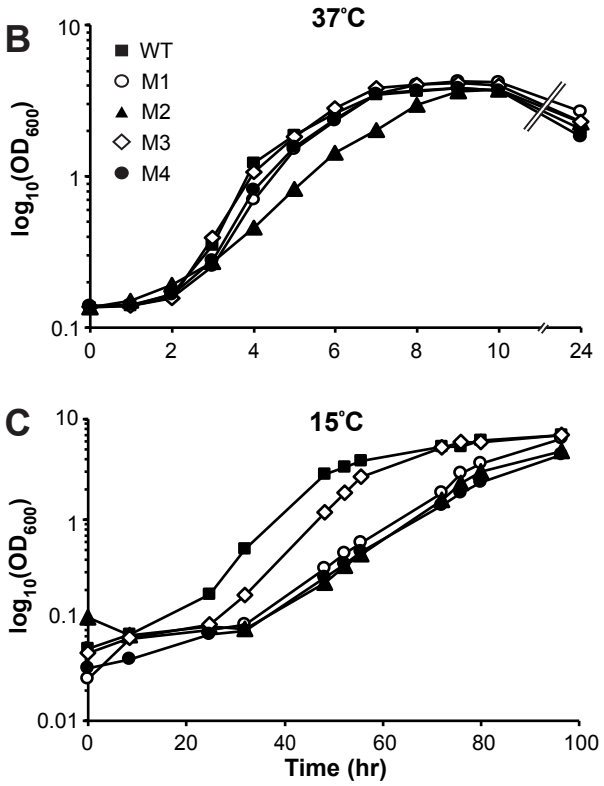
nuclease probing analyses suggest that Helix H3 is not involved in L20-binding interactions (Choonee et al. 2007), and this stem is only present in ~75% of examples of this RNA (Deiorio-Haggar et al. 2013). As anticipated, regulation was retained for the M3 mutant construct (~5-fold repression) (Figure 3.1B,C).

Finally, the M4 mutation is designed to destabilize the stem of the Rho-independent terminator that forms upon L20 binding (Figure 3.1A). This mutation resulted in elevated constitutive expression. The activity measured for the M4 mutant under both unrepressed (empty plasmid) and repressed (*infC* operon overexpression) conditions was substantially higher than that obtained with the wild-type RNA (Figure 3.1B,C). Regulatory activity with the M4 mutant RNA was significantly reduced (~2.5-fold repression upon *infC* operon overexpression).

3.2.2 L20-interacting RNA mutant recombinant strains are cold-sensitive

To investigate the effects mutations to the L20-interacting RNA and the loss of *infC* ribosomal protein operon regulation have on *B. subtilis* fitness, we replaced the native copy of the L20-interacting RNA within the *B. subtilis* 168 genome with either a wild-type or mutant recombinant version via homologous recombination (Figure 3.2A). Growth curves were performed to measure recombinant strain fitness, as translation efficiency and consequently ribosome quality and quantity can be inferred from log phase growth (Harvey 1970; Scott et al. 2014). Strains were assayed in nutrient-rich 2XYT medium at both 37ºC and 15ºC because sensitivity to low temperatures can be indicative of defects in ribosome assembly, composition, and/or function (Guthrie et al. 1969; Tai et al. 1969; Feunteun et al. 1974a; Isono and Krauss 1976; Isono et al. 1977; Charollais et al. 2004; Bharat and Brown 2014; Choudhury and Flower 2015).





D

Doubling Time (min)

Strain	37ºC	15ºC
168	52 ± 1	346 ± 3
WT	68 ± 2	385 ± 3
M1	69 ± 1 (1.0)	538 ± 31 (1.4)*
M2	91 ± 4 (1.3)*	556 ± 28 (1.4)*
M3	60 ± 1 (0.9)†	379 ± 4 (1.0)
M4	68 ± 3 (1.0)	546 ± 22 (1.4)*

Figure 3.2. Construction and growth of L20-interacting RNA native locus recombinant strains. (A) L20-interacting RNA recombinant strain design. The second *infC* operon promoter (P_{infC2}), L20-interacting regulatory RNA sequence, and two ~500 bp regions flanking the promoter and regulatory RNA locus were PCR-amplified from *B. subtilis* 168 genomic DNA. A PCR product in which an erythromycin resistance cassette (*erm*) was introduced immediately upstream of the second *infC* operon promoter was generated, sub-cloned, and transformed into *B. subtilis* 168. Integration of the PCR constructs at the *infC* locus via double-crossover homologous recombination replaced the native L20-interacting RNA with a wild-type or mutant recombinant version. Growth curves for each recombinant strain in 2XYT at 37ºC (B) and 15ºC (C). Growth assays were performed three or more times for each strain. Representative curves are shown. (D) Doubling times (min) of L20-interacting RNA recombinant strains grown in 2XYT at 37ºC and 15ºC. Values were calculated from log phase OD_{600} values and are the mean of three or more independent experimental replicates; ± indicates the standard error of the mean across biological replicates. Numbers in parentheses denote mutant recombinant strain doubling time relative to that of the wild-type recombinant (WT) strain at the corresponding temperature. Asterisks (*) indicate mutant recombinant strains that grew significantly slower than the wild-type recombinant at the corresponding temperature. Daggers (†) indicate mutant recombinant strains that grew significantly faster than the wild-type recombinant at the corresponding temperature ($p<0.05$). 168, parental strain.

The M1, M3, and M4 mutant recombinant strains grew similarly to, if not better than, the wild-type recombinant strain at 37ºC, with doubling times of 69, 60, and 68 minutes, respectively, in comparison to a doubling time of 68 minutes for the wild-type recombinant strain (Figure 3.2B,D). The M2

protein-binding mutant recombinant strain grew approximately 1.3 times slower than the wild-type recombinant at 37ºC, with a doubling time of 91 minutes. The M2 regulatory mutant recombinant strain was consistently unstable and difficult to maintain. While it is possible that additional mutations elsewhere in the genome contribute to the instability and slow growth exhibited by this strain, this growth defect at 37ºC was observed for multiple M2 mutant recombinant strains generated from independent transformation events.

In contrast, all three mutant recombinant strains in which the regulatory activity of the L20-interacting RNA was compromised (M1, M2, M4) grew approximately 1.4 times slower than the wild-type recombinant at 15ºC, whereas the M3 control mutant recombinant strain grew comparably to the wild-type recombinant at this temperature (Figure 3.2C,D). This significant cold-temperature sensitive growth defect suggests that improper *infC* operon regulation likely affects ribosome production.

The 168 parental strain did not demonstrate any cold temperature-sensitive growth defects; rather, it consistently grew faster than the wild-type recombinant strain at both 37ºC and 15ºC. This difference is likely due to the presence of the erythromycin resistance cassette within the wild-type recombinant strain and/or the use of antibiotic in the recombinant strain medium. Consequently, the doubling times of the 168 parental strain are included in Figure 3.2D for reference, but the strain is not shown in the representative growth curves.

3.2.3 Position of erythromycin resistance cassette does not influence recombinant strain phenotype

Following construction of our recombinant strains, we noticed that we failed to properly incorporate the first *infC* operon promoter into our native locus recombinant

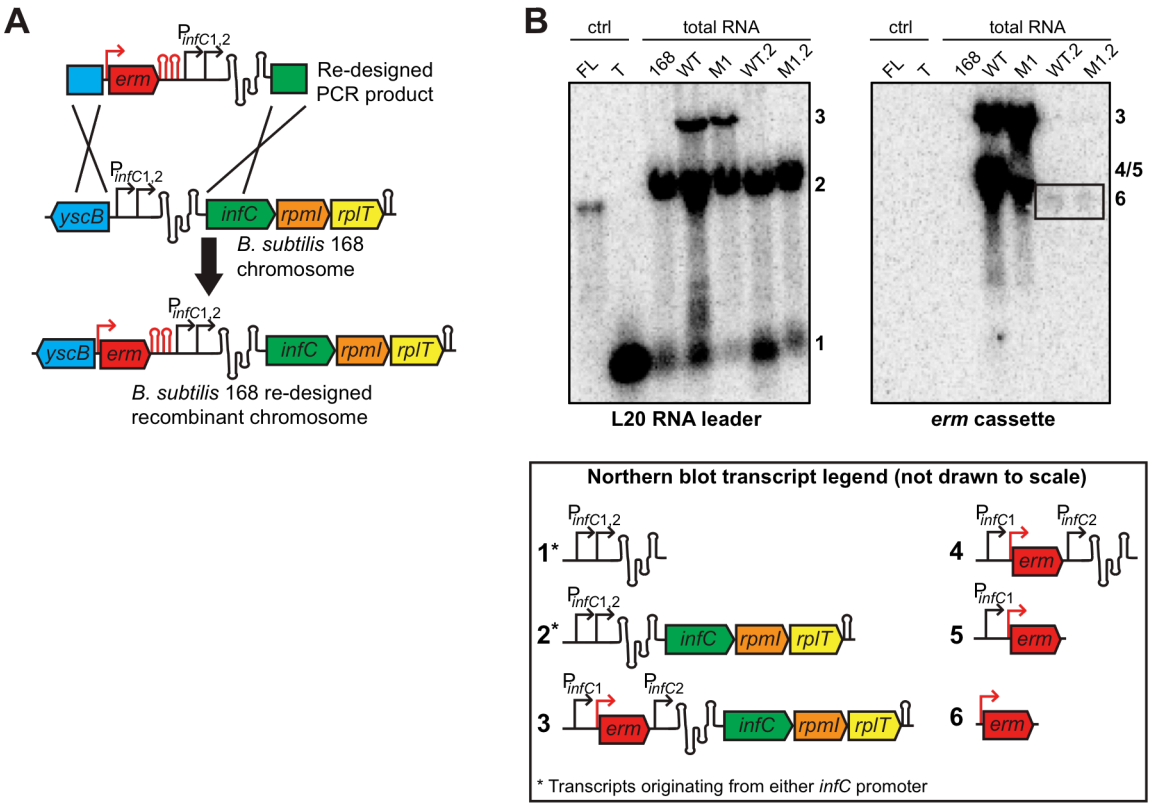


Figure 3.3. Construction of re-designed L20-interacting RNA recombinant strains. (A) Re-designed L20-interacting RNA recombinant strain constructs. New PCR products were constructed to preserve the native organization of the two *infC* operon promoters (P_{infc1} and P_{infc2}). Re-designed recombinant strains were generated as described in Figure 3.2A, except the *erm* cassette was introduced into the intergenic region immediately upstream of the first *infC* operon promoter and two transcription terminators were appended onto the 3' end of the *erm* cassette to prevent potential read-through from the *erm* promoter (red arrow). (B) Northern blot analysis of total log phase RNA from original (WT, M1) and re-designed (WT.2, M1.2) recombinant L20-interacting RNA strains grown in 2XYT at 37ºC. The membrane was hybridized with oligonucleotide probes complementary to the L20-interacting RNA leader (left) and the erythromycin (*erm*) resistance cassette (right) (Table 3.1). *In vitro* transcribed RNAs corresponding to the full length (ctrl FL) and terminated (ctrl T) *infC-rpmI-rplT* operon transcripts were included as size standards. Transcripts 4 and 5 (see Northern blot transcript legend) are potential transcripts that cannot be distinguished on the blots shown; transcript 4 may also be present in the original recombinant strains (WT, M1) on the L20 RNA leader blot (left). 168, parental strain.

strain design. In *B. subtilis*, the *infC* operon is under the control of two promoters. Transcription from the first promoter adds 58 nucleotides to the 5' end of the *infC* operon transcript, which contains an RNase Y cleavage site that ultimately allows for increased expression of ribosomal proteins L35 and L20 relative to IF3 (Bruscella et al. 2011). In

our original recombinant strain design, we introduced the erythromycin resistance (*erm*) cassette immediately upstream of the second *infC* operon promoter, displacing the position of the first *infC* operon promoter (Figure 3.2A). Consequently, Northern blots revealed extra transcript originating from read-through from the displaced promoter and/or the *erm* cassette promoter in the recombinant strains, and this was confirmed with 5'-RACE (Figure 3.3B, Northern blot transcript 3).

To determine if disrupting the context of the two promoters and the resulting expression of the extra transcript influenced strain phenotype, we re-designed our wild-type recombinant and M1 mutant recombinant strains to more accurately reflect the native organization of the *infC* operon promoter region within the *B. subtilis* genome. In the re-designed recombinant strains, we appended two terminators onto the 3' end of the *erm* cassette coding sequence to prevent potential read-through from the *erm* promoter and inserted the *erm* cassette into the intergenic region immediately upstream of the first *infC* operon promoter (Figure 3.3A). The extra transcript previously observed in our original recombinant strains was not detected in our re-designed recombinant strains, suggesting that the native organization of the *infC* operon promoter region was restored (Figure 3.3B, Northern blot transcript 3). Consistent with our previous growth assay results, the re-designed M1 mutant recombinant strain (M1.2) exhibited cold-sensitive growth defects similar to that of the original M1 regulatory mutant recombinant strain (Figure 3.4). Therefore, we continued to use the original recombinant strains we constructed for all subsequent experiments.

It should be noted that although similar trends in growth were observed for both recombinant strain designs during the growth assays, the re-designed recombinant strains grew slower than the original recombinant strains at both 37ºC and 15ºC. This is likely due to differences in *erm* cassette expression. Transcription from the displaced

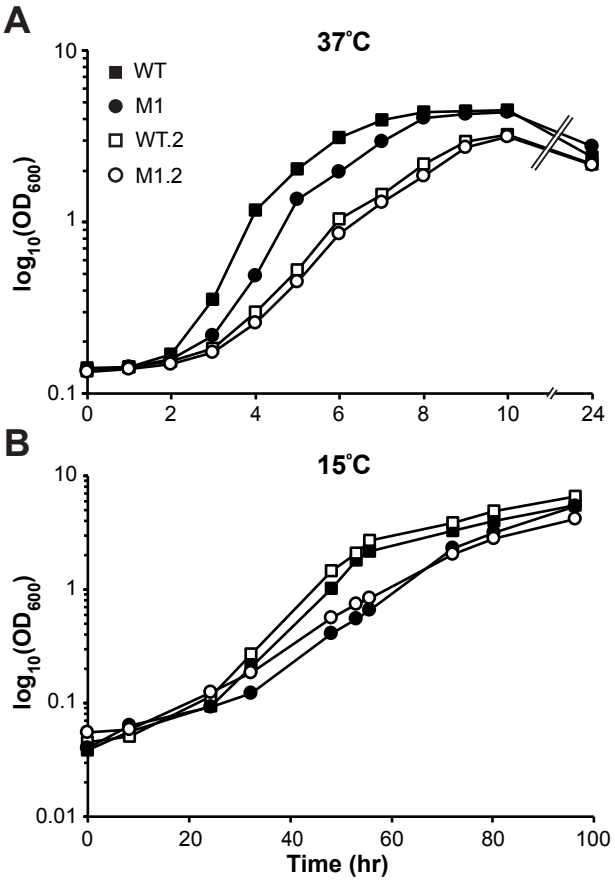


Figure 3.4. Comparison of original (WT, M1) and re-designed (WT.2, M1.2) L20-interacting RNA recombinant strain growth. Growth curves for each recombinant strain grown in 2XYT at 37ºC (A) and 15ºC (B). Growth assays were performed two or more times for each strain. Representative curves are shown. (C) Doubling times (min) of original and re-designed L20-interacting RNA recombinant strains grown in 2XYT at 37ºC and 15ºC. Values were calculated from log phase OD_{600} values and are the mean of two or more independent experimental replicates; ± indicates the standard error of the mean across biological replicates. Numbers in parentheses denote mutant recombinant strain doubling time relative to that of the appropriate wild-type recombinant (WT or WT.2) strain at the corresponding temperature. Symbols (*,^,#) indicate mutant recombinant strains that grew significantly slower than the appropriate wild-type recombinant at the corresponding temperature; the p values are indicated in the figure. 168, parental strain.

C

Doubling Time (min)

Strain	37ºC	15ºC
168	53 ± 1	346 ± 4
WT	68 ± 2	385 ± 3
M1	69 ± 1 (1.0)	538 ± 31 (1.4)^
WT.2	86 ± 1	414 ± 22
M1.2	89 ± 1 (1.0)*	720 ± 23 (1.7)#

* p = 0.048
^ p = 0.019
# p = 0.005

first *infC* operon promoter greatly increased expression of the *erm* cassette in the original recombinant strains relative to that of the re-designed recombinant strains (Figure 3.3B, Northern blot transcripts 4-6). All growth assays were performed in medium containing a final concentration of 0.5 μg/mL erythromycin; thus, the slower doubling times calculated for the re-designed recombinant strains may be due to an increased sensitivity to erythromycin.

3.2.4 Low temperatures exacerbate mutant recombinant strain *infC* operon mis-regulation at stationary phase

To further assess the origins of the temperature-sensitive phenotype observed during the growth assays, we quantified levels of the native *infC* operon transcript in the recombinant strains during both log (OD_{600} ~0.3-0.7) and early stationary phase (OD_{600} ~2.0-3.0) at 37ºC and 15ºC in the presence of endogenous L20 protein only. Because the L20-interacting RNA utilizes a transcription termination mechanism to regulate *infC* operon expression, we measured transcript levels using quantitative RT-PCR (qRT-PCR) with primers targeting the *rplT* (L20) coding region. While these measurements cannot capture the instantaneous rates of transcription or translation, the relative levels of *rplT* transcript do give some indication of whether differences in termination occur under these conditions, and whether there are likely to be differences in protein levels.

The M2 and M3 mutant recombinant strains exhibited *rplT* transcript levels comparable to that of the wild-type recombinant during log phase at 37ºC, while the M1 and M4 mutant recombinant strains demonstrated a ~2.7 and 1.7-fold increase in *rplT* levels, respectively, in comparison to that of the wild-type recombinant strain under these conditions (Figure 3.5A). During early stationary phase at 37ºC, *rplT* transcript levels in all strains were approximately 2-fold higher relative to the *nifU* internal control compared to log phase, and the increased transcript levels measured in the M1 and M4 mutant recombinant strains were even more prominent (~4.3 and 5-fold increase in *rplT* levels, respectively, compared to the wild-type recombinant). Under these conditions, *rplT* levels in the M2 mutant recombinant strain remained consistent with those in the wild-type recombinant, and the M3 control mutant recombinant strain displayed a ~2.5-fold decrease in *rplT* levels relative to the wild-type recombinant strain. The β-galactosidase activities measured for each L20-interacting RNA reporter in the

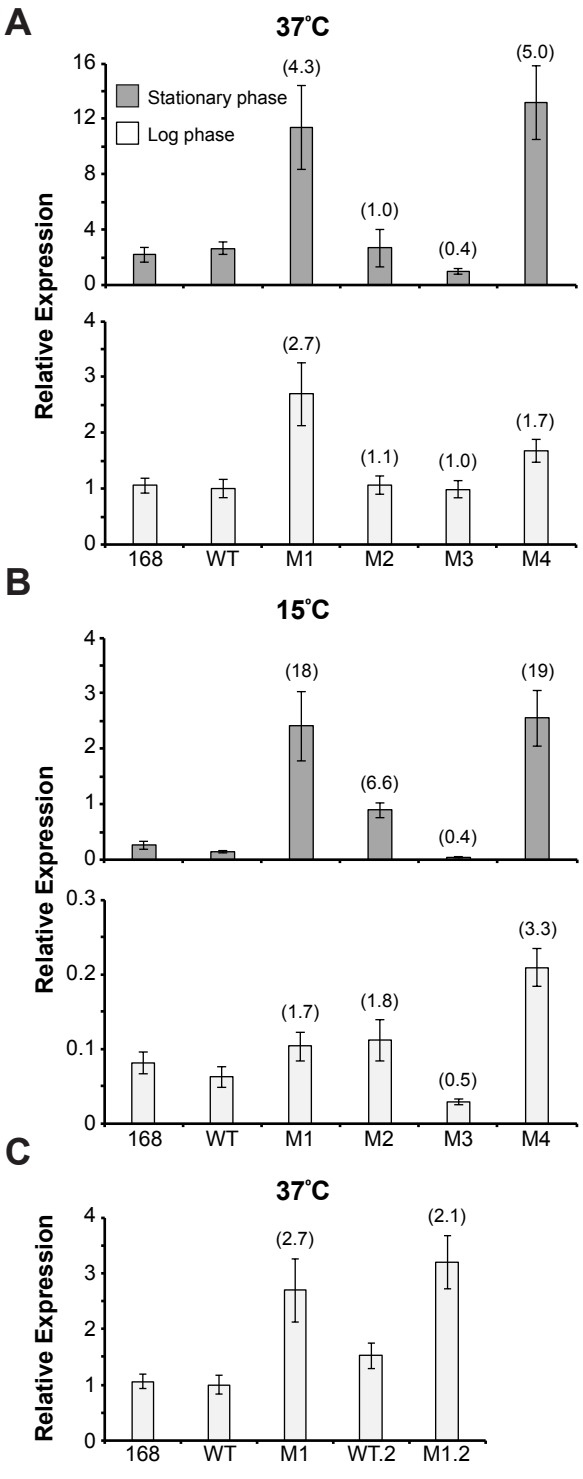


Figure 3.5. Recombinant strain *infC* operon transcript levels. qRT-PCR quantification of the native *rplT* transcript in each of the recombinant strains grown to log and early stationary phase at 37ºC (A) and 15ºC (B). For each strain/condition, *rplT* transcript level was normalized to the level of housekeeping gene *nifU* (Reiter et al. 2011). Graphs depict relative *rplT* transcript levels in each strain compared to the *rplT* level in the wild-type recombinant (WT) strain grown to log phase at 37ºC. Error bars represent the standard error of the mean across three technical replicates. Numbers in parentheses denote mutant recombinant strain *rplT* transcript level relative to that of the wild-type recombinant strain at the corresponding growth phase and temperature. (C) qRT-PCR quantification of the native *rplT* transcript in the re-designed recombinant strains (WT.2, M1.2) grown to log phase at 37ºC, normalized and graphed as described above. For comparison, data from A for the 168 parental and original WT and M1 recombinant strains are included. Numbers in parentheses denote mutant recombinant strain *rplT* transcript levels relative to that of the corresponding wild-type recombinant strain. Consistent with the qRT-PCR results from the original recombinant strains, the M1.2 recombinant strain exhibited elevated *rplT* levels relative to that of the WT.2 recombinant strain.

absence of *infC* operon overexpression (empty plasmid) largely reflect the relative transcript levels measured at 37ºC for each native locus recombinant strain using qRT-PCR (Figure 3.1B). Overall, the differences in *rplT* transcript levels measured at 37ºC do not appear to influence or correlate with strain growth at this temperature.

At 15ºC, *rplT* transcript levels decreased relative to the *nifU* internal control for all strains during both log and early stationary phase in comparison to the transcript levels quantified at 37ºC (Figure 3.5B). This is expected, as production of ribosomal components decreases during periods of limited growth and/or under poor growth conditions (Nomura et al. 1984; Bremer and Dennis 1996; Maguire 2009). However, all three regulatory mutant recombinant strains (M1, M2, M4) exhibited a ~2 to 3-fold increase in *rplT* levels during log phase at 15ºC relative to that of the wild-type recombinant strain, and these elevated transcript levels were considerably more pronounced during early stationary phase at this temperature. Similar to the transcript levels quantified at 37ºC, *rplT* transcript in the M3 control mutant recombinant strain was approximately half of that measured in the wild-type recombinant at 15ºC. The qRT-PCR data at 15ºC suggest that the compromised negative-feedback regulation in the M1, M2, and M4 mutant recombinant strains results in unnecessary constitutive expression of the *infC* ribosomal protein operon. This expression likely impacts strain growth when rapid ribosome synthesis is not required, such as during entry into stationary phase and the cold temperature conditions assessed in this study.

3.2.5 L20-interacting RNA mutants demonstrate improper rRNA processing at low temperatures

The cold-sensitive phenotype of the L20-interacting RNA regulatory mutant recombinant strains suggests that removing the RNA-based regulation of the *infC* operon results in aberrant ribosome assembly. To examine if our regulatory mutant recombinant strains are defective in ribosome biosynthesis, we analyzed rRNA 5' end processing in each strain grown to log phase at both 37ºC and 15ºC using primer extension. Ribosome assembly and rRNA maturation are tightly coupled. Pre-rRNA

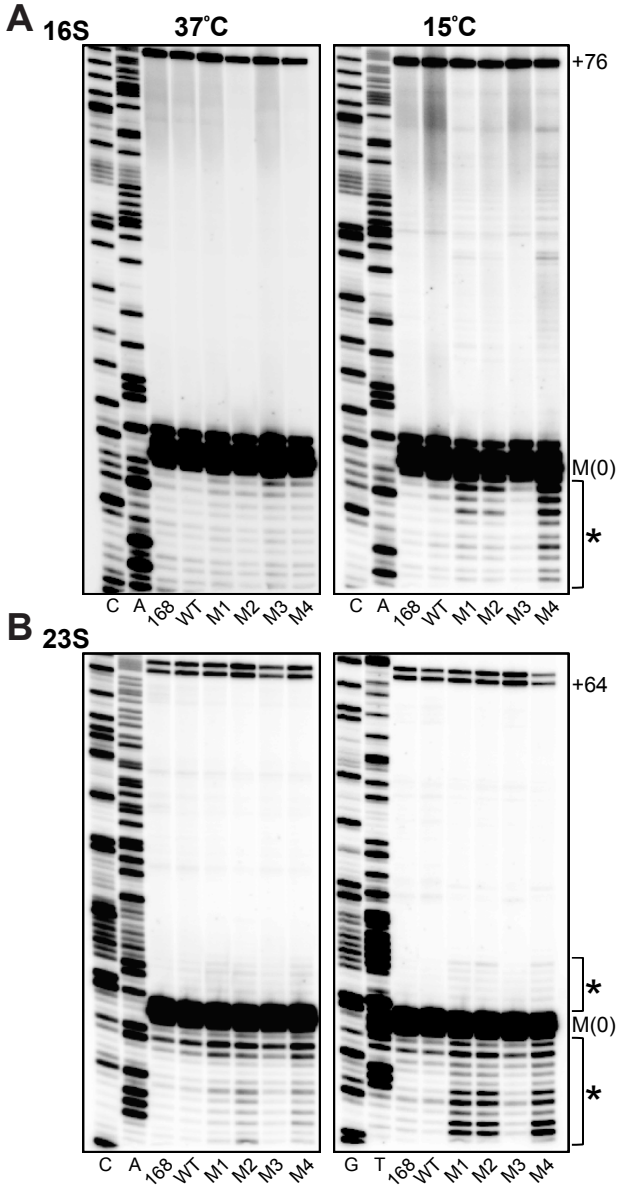


Figure 3.6 L20-interacting RNA recombinant strain rRNA processing. Primer extension analysis of 16S rRNA (A) and 23S rRNA (B) from L20-interacting RNA recombinant strains grown to log phase in 2XYT at 37ºC and 15ºC. Sequencing reactions performed with the same primers and *in vitro* transcribed RNAs corresponding to portions of the unprocessed 5' ends of both the 16S and 23S rRNAs were used to identify the mature 5' end of each rRNA, labeled as M(0). Sequencing reactions are labeled as their complements to allow for direct reading. Previously characterized precursor species are labeled according to the number of extra nucleotides relative to M(0) and are included for reference (+76 for 16S rRNA, +64 for 23S rRNA) (Britton et al. 2007; Redko et al. 2008). Asterisks (*) indicate regions in which mutant recombinant strain rRNA processing differs from that of the wild-type recombinant (WT) strain. Primer extension analysis was conducted with RNA extracted from three or more independent biological replicates of each strain grown to log phase at both 37ºC and 15ºC. Representative gels are shown. 168, parental strain.

transcripts are cleaved to their mature forms concurrent with ribosomal subunit assembly; thus, accumulation of precursor rRNA transcripts and/or improper rRNA processing can be indicative of defects in ribosome composition (Feunteun et al. 1974b; Charollais et al. 2003, 2004; Jain 2008; Choudhury and Flower 2015).

Consistent with our previous findings, little to no differences in the 5' end processing of both the 16S and 23S rRNAs were observed for all recombinant strains at 37ºC (Figure 3.6). The multiple species of closely spaced mature 5' ends of each rRNA (indicated as M(0)) have been previously characterized and are due to differences in

rRNA sequence across the 10 *rrn* operons encoded within the *B. subtilis* 168 genome, different RNase cleavage sites, and/or other auxiliary processing pathways (Srivastava and Schlessinger 1990; Britton et al. 2007; Redko et al. 2008).

When the strains were grown at 15ºC, multiple extension products longer and shorter than those corresponding to the mature 5' ends of the 16S and 23S rRNAs were apparent in the primer extension reactions of the mutant recombinant strains in which *infC* operon regulation was compromised (M1, M2, M4) (Figure 3.6). While the amount of extension product corresponding to the mature 5' ends of both rRNAs (M(0)) was comparable to those of the 168 parental, wild-type recombinant, and M3 control mutant recombinant strains at this temperature, the extra extension species are likely due to improper rRNA processing brought about by aberrant ribosome assembly or an increase in rRNA degradation as a result of the cold-sensitive growth defects previously noted. Cold-sensitive rRNA processing defects have also been observed in bacterial strains lacking specific ribosomal assembly factors (Charollais et al. 2003, 2004; Jain 2008; Bharat and Brown 2014; Choudhury and Flower 2015). None of these additional extension products were present in the 16S and 23S rRNA primer extension reactions of the 168 parental, wild-type recombinant, and M3 control mutant recombinant strains at this temperature. These results are consistent with those from our growth assays and further suggest that disrupting the RNA-based negative-feedback regulation of the *infC* operon leads to defects in ribosome synthesis or assembly.

3.2.6 Mutations to the L20-interacting RNA affect ribosomal subunit sedimentation

To further characterize the impact disruptions to *infC* operon regulation have on ribosome assembly, we analyzed ribosomal subunit sedimentation profiles of the mutant

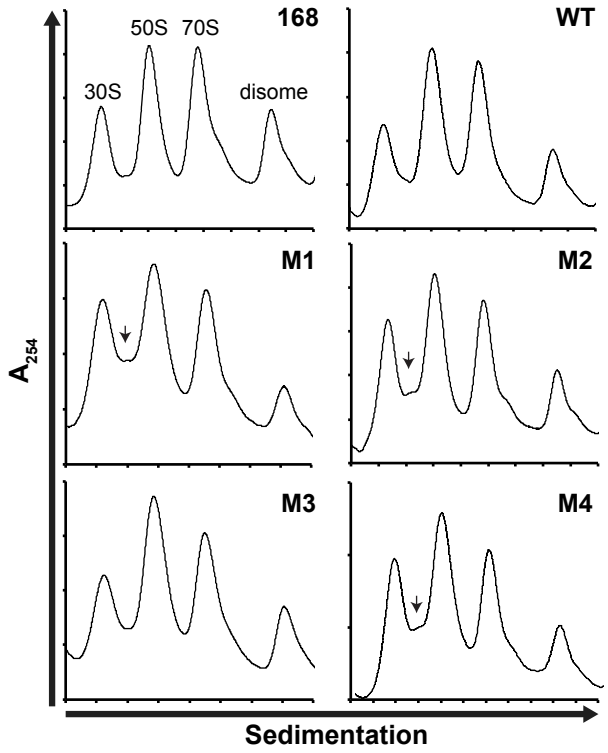


Figure 3.7. Ribosome sedimentation profiles of L20-interacting RNA recombinant strains grown at 15ºC. Cells were grown in 2XYT at 15ºC with shaking (225 rpm) until an OD_{600} ~0.3-0.4 was reached. Ribosomal subunit sedimentation profiles were resolved by 10-55% (w/v) sucrose density gradients. Arrows indicate peaks that are present in the profiles of the regulatory mutant recombinant strains, but not in those of the control strains.

recombinant strains using sucrose density gradient ultracentrifugation. Because defects in rRNA processing were only apparent in the regulatory mutant recombinant strains during growth at cold temperatures, sedimentation profiles were generated from extracts of cells grown to early log phase at 15ºC. For all strains, the peak corresponding to complete 70S ribosomes was reduced relative to that of the 30S and 50S subunits (Figure 3.7). This likely reflects a global reduction in active translation due to suboptimal growth at 15ºC (Dai et al. 2016). Although the ratio of 30S subunits to 50S subunits remained fairly consistent across all strains, a small peak was observed between the 30S and 50S peaks in the sedimentation profiles of the mutant recombinant strains in which the regulatory activity of the L20-interacting RNA was compromised (M1, M2, M4). This peak suggests an accumulation of precursor 50S subunit particles in the regulatory mutant recombinant strains and is indicative of defects in ribosome biosynthesis. Similar peaks corresponding to 50S subunit intermediates have been observed in the sedimentation profiles of bacterial strains in which ribosomal assembly factors have been deleted (Charollais et al. 2003, 2004; Jain 2008) and/or select ribosomal proteins have been overexpressed (Sykes et al. 2010). These data, in combination with the

results from our rRNA 5' end processing assays, indicate that the impact of unregulated *rplT* expression is the formation or accumulation of improper ribosome assembly products.

3.3 Discussion

Many studies have identified, characterized, and compared the structures and mechanisms of the RNA elements responsible for the regulation of ribosomal proteins in bacteria. Furthermore, as the amount of available genomic data continues to increase, additional ribosomal protein *cis*-regulatory RNAs are being discovered, especially in non-model species of bacteria. Many of these newly identified regulatory RNA structures are distinct from those previously characterized and are narrowly distributed to select bacterial phyla or classes. Despite these advances in the field, our understanding of the factors that influence the formation, conservation, and adaptive evolution of these diverse RNA *cis*-regulators within bacterial genomes is relatively nonexistent. To address these questions, we investigated how mutations to the native locus of the L20-interacting RNA *cis*-regulator of ribosomal protein synthesis within the *B. subtilis* genome impact cell fitness. We show that disrupting the regulatory activity of this RNA structure results in constitutive downstream expression, reduced growth, aberrant rRNA processing, and the accumulation of a 50S ribosomal precursor species at low temperatures.

Differences in *rplT* transcript levels as measured by qRT-PCR due to mutation and/or mis-regulation of the *infC* operon do not appear to significantly impact recombinant strain phenotype at 37ºC. The M1 and M4 regulatory mutant recombinant strains exhibited an increase in *rplT* levels at this temperature, but both strains grew comparably to the wild-type recombinant at 37ºC. Although the M2 regulatory mutant

recombinant strain demonstrated *rplT* transcript levels similar to those of the wild-type recombinant during both log and early stationary phase at 37ºC, the strain grew approximately 1.3 times slower than the wild-type recombinant strain at this temperature. Furthermore, despite these differences in *rplT* transcript levels and growth, all three regulatory mutant recombinant strains exhibited rRNA processing comparable to that of the wild-type recombinant strain at 37ºC.

In contrast, *rplT* expression resulting from the loss of *infC* operon regulation had more severe impacts on regulatory mutant recombinant strain phenotype at 15ºC. The M1, M2, and M4 mutant recombinant strains all exhibited elevated constitutive *rplT* transcript levels in comparison to that of the wild-type recombinant at 15ºC, and the increased transcript levels were more pronounced during early stationary phase. Subsequently, all regulatory mutant recombinant strains demonstrated similar temperature-sensitive growth, rRNA processing, and ribosome assembly product distribution defects at 15ºC.

Unlike the mutations that disrupt the function of the L20-interacting RNA regulator, the M3 control mutation did not cause any mutant phenotypes. Our β-galactosidase assays indicate that the M3 reporter construct retained regulatory activity and the M3 mutant recombinant strain showed no cold-sensitive growth defects and no changes in rRNA processing and ribosomal subunit distribution at 15ºC compared to the 168 parental and wild-type recombinant strains. Although *rplT* transcript levels in the M3 mutant recombinant strain were lower than that of the wild-type recombinant at both 37ºC and 15ºC (especially during early stationary phase), the lack of mutant phenotype at both temperatures suggests that this under-expression does not significantly impact the biological outcome in these conditions. The behavior of the M3 control mutant recombinant strain further indicates that the cold-sensitive defects of the L20-interacting

RNA M1, M2, and M4 regulatory mutant recombinant strains are primarily the result of unnecessary constitutive expression due to compromised *infC* operon autoregulation. Taken together, these results suggest that RNA *cis*-regulators of ribosomal protein synthesis are important for maintaining a balance between ribosomal protein operon expression and repression, especially under conditions in which the production of ribosomal components is not required, such as during entry into stationary phase or periods of slow or suboptimal growth at low temperatures.

Sensitivity to cold temperatures, reduced growth, rRNA maturation defects, and/or aberrant ribosomal subunit sedimentation profiles are hallmark characteristics of bacterial strains that harbor ribosomal protein mutations or deletions (Guthrie et al. 1969; Tai et al. 1969; Feunteun et al. 1974b; Isono and Krauss 1976; Isono et al. 1977). These phenotypes have also been observed in bacteria defective in specific rRNA helicases, rRNA or ribosomal protein modification enzymes, rRNA operon copy number, ribosome assembly GTPases, and other assembly and translation cofactors (Charollais et al. 2003, 2004; Jain 2008; Bharat and Brown 2014; Choudhury and Flower 2015; Gyorfy et al. 2015). In this work, we demonstrate that constitutive expression driven by loss of autoregulation from subtle point mutations to a *cis*-regulatory RNA structure within the 5'-UTR of a ribosomal protein operon transcript has similar impacts on ribosome assembly and cell fitness. Our findings reinforce the importance of the coordinated and stoichiometric production of ribosomal components for proper ribosome biosynthesis in bacteria. While the function of RNA *cis*-regulators of ribosomal protein synthesis is well understood, our study highlights the role of these regulatory RNA elements in optimal cell growth and informs us on the factors that influence the formation, evolution, and conservation of structured RNA regulators within bacterial genomes. Understanding the fitness costs associated with the loss of ribosomal protein RNA *cis*-regulators sets the

stage for the development of novel antimicrobials that target ribosome synthesis and assembly.

3.4 Materials and Methods

3.4.1 β-galactosidase activity assay plasmid and strain construction

To generate the protein overexpression plasmid and combat against the growth defects previously noted with strains that solely overexpressed L20 (Choonee et al. 2007), the complete *infC-rpmI-rplT* operon, including the native *infC* Shine-Dalgarno sequence, was PCR-amplified from *B. subtilis* 168 genomic DNA (GenBank: AL009126, complement of 2952224-2953363) with primers containing EcoRI and PstI restriction sites and changing the *infC* ATT start codon to ATG for stronger expression (Table 3.1). After digestion, the PCR product was cloned into the pYH213 plasmid digested with the same enzymes, upstream from a P_{T7A1}-*lacO* IPTG-inducible promoter (Yakhnin et al. 2015). This plasmid, as well as an empty control plasmid (pAY132) (Yakhnin et al. 2015), were transformed into *B. subtilis* 168 as described previously (Yasbin et al. 1975). Transformants were screened for resistance to tetracycline (12.5 μg/mL) and verified via PCR.

To generate the *infC'-'lacZ* reporter constructs, the region containing the second native *infC* operon promoter (the major promoter for *infC* expression (Choonee et al. 2007; Bruscella et al. 2011)), the wild-type L20-interacting RNA leader, and Shine-Dalgarno sequence and first nine codons of *infC* was PCR-amplified from *B. subtilis* 168 genomic DNA (GenBank: AL009126, complement of 2953323-2953586) with primers containing EcoRI and BamHI restriction sites and changing the *infC* ATT start codon to ATG for stronger expression (Table 3.1). The PCR product was cloned in-frame as a

translational fusion with a *lacZ* reporter into a modified pDG1728 plasmid digested with the same enzymes (Guérout-Fleury et al. 1996; Babina et al. 2017). Mutations to the L20-interacting RNA were obtained by site-directed mutagenesis or PCR assembly and verified via Sanger sequencing (Table 3.1). Reporter constructs were transformed into the above *B. subtilis* 168 pYH213/pAY132 protein overexpression strains as described previously (Jarmer et al. 2002). Transformants were selected on TBAB + 12.5 μg/mL tetracycline + 100 μg/mL spectinomycin and screened for proper integration of the *lacZ* reporter constructs into the *amyE* locus based on sensitivity to erythromycin (0.5 μg/mL) and loss of amylase activity (plating on TBAB + 1% starch and staining with Gram's iodine solution, Sigma-Aldrich).

3.4.2 β-galactosidase activity assays

B. subtilis 168 *lacZ* reporter strains were grown from single colonies in 2 mL 2XYT + 12.5 μg/mL tetracycline + 100 μg/mL spectinomycin for 16-18 hours at 37ºC with shaking (225 rpm). These cultures (30 μl for the pAY132 strains, 50 μl for the pYH213-*infC-rpmI-rplT* strains) were used to inoculate 2 mL 2XYT + 12.5 μg/mL tetracycline + 100 μg/mL spectinomycin + 1 mM IPTG cultures, which were then grown at 37ºC with shaking (225 rpm) until an OD_{600} ~0.3-0.7 was reached. Cells (1 mL) were harvested and re-suspended in 1 mL Z buffer (50 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$, 50 mM 2-Mercaptoethanol) + 100 μg/mL spectinomycin. β-galactosidase activity assays were performed as previously described using 0.05 mL of cell suspensions. Miller Units were calculated as follows (Miller 1992):

$$\text{Miller Units} = 1000 * \frac{A_{420}}{\Delta t\ (\text{min.}) * A_{600} * vol.(mL)}$$

The values reported represent the mean of three or more independent replicates; error bars represent standard error of the mean across biological replicates. To determine the fold repression for each RNA reporter construct, the mean Miller Units for each empty plasmid reporter strain was divided by that of the corresponding *infC* operon overexpression strain. Standard error for the fold repression values were propagated from the error calculated for the Miller Units as described previously (Taylor 1997).

3.4.3 L20-interacting RNA native locus recombinant strain construction

The recombinant strains for the growth assays, qRT-PCR, rRNA processing assays, and ribosome sedimentation profiles were generated as described previously (Babina et al. 2017). Briefly, the second *infC* operon promoter, wild-type L20-interacting RNA leader, and two ~500 bp regions of homology flanking the promoter and regulatory RNA leader region were PCR-amplified from *B. subtilis* 168 genomic DNA (GenBank: AL009126; complement of 2953587-2954227 for the 5' flanking ~500 bp region of homology, complement of 2953142-2953586 for the region containing the promoter, RNA leader, and 3' flanking ~500 bp region of homology). A PCR product in which an erythromycin resistance cassette was introduced into the intergenic region immediately upstream from the second *infC* operon promoter was generated and cloned into pCR2.1 or pCR4 TOPO-TA vector (Invitrogen). Mutations to the L20-interacting RNA were obtained via site-directed mutagenesis or PCR assembly (Table 3.1). These constructs were then transformed into *B. subtilis* 168 as described previously and transformants were screened for resistance to erythromycin (0.5 μg/mL) (Jarmer et al. 2002). Integration of the complete recombinant construct within the *infC* locus and the presence of the L20-interacting RNA mutations of interest were verified via PCR and Sanger sequencing.

To generate the re-designed recombinant L20-interacting RNA strains in which the first native *infC* promoter immediately precedes the second *infC* promoter, new 5' and 3' ~500 bp regions of homology that flank the *infC* locus were PCR-amplified from *B. subtilis* 168 genomic DNA (GenBank: AL009126; complement of 2953643-2954227 for the 5' flanking ~500 bp region of homology, complement of 2953142-2953642 for the region containing the promoters, regulatory RNA, and 3' flanking ~500 bp region of homology). To prevent read-through from the erythromycin resistance cassette promoter, the region containing two Rho-independent terminators was PCR-amplified from the pYH213 plasmid (Yakhnin et al. 2015) and appended onto the 3' end of the erythromycin resistance cassette via PCR assembly. Recombinant constructs were assembled and transformed into *B. subtilis* 168 as described above.

3.4.4 Northern blots

Total RNA was extracted from select *B. subtilis* 168 strains grown to log phase (OD_{600} ~0.3-0.7) in 20 mL 2XYT (+ 0.5 μg/mL erythromycin for recombinant strains) at 37ºC with shaking (225 rpm). As a size-standard, template DNA corresponding to both the full-length and terminated *infC-rpmI-rplT* operon transcripts were PCR-amplified from *B. subtilis* 168 genomic DNA using forward primers that contained the T7 promoter sequence (GenBank: AL009126; complement of 2952168-2953550 for the full-length transcript, complement of 2953394-2953550 for the terminated transcript) (Table 3.1). T7 RNA polymerase was used to transcribe RNA from these templates, and transcription reactions were ethanol precipitated and re-suspended in TE buffer (Milligan et al. 1987). These transcripts (15 ng of full-length transcript, 3 ng of terminated transcript) and 10 μg of each total RNA sample were separated on a 0.8% denaturing agarose gel and transferred to an Amersham Hybond-N+ membrane (GE Healthcare) overnight, as

described previously (Rio 2015). The membrane was UV-crosslinked and pre-hybridized in 10 mL Ambion ULTRAhyb ultrasensitive hybridization buffer (Life Technologies) at 37ºC or 42ºC for ~2-4 hours with rotation. Synthetic oligonucleotide DNA probes (40 pmol, IDT) complementary to transcript regions of interest were 5'-end labeled with [γ-^{32}P] ATP (Perkin Elmer) (Regulski and Breaker 2008) and allowed to hybridize to the membrane overnight at 37ºC or 42ºC with rotation (Table 3.1). The membrane was washed twice with approximately 100 mL 2X SSC (300 mM NaCl, 30 mM trisodium citrate [pH 7.0]) + 0.1% SDS at 37ºC or 42ºC for 20 minutes with rotation, exposed to a phosphor screen for 16-72 hours, and imaged using a Typhoon FLA 9500 scanner (GE Life Sciences). The membrane was then stripped with boiling hot water + 0.1% SDS at 70ºC with rotation until signal was removed and re-probed for a different transcript as described above.

3.4.5 5'-RACE

Total RNA was extracted from a log phase (original) wild-type recombinant *B. subtilis* 168 culture grown at 37ºC as described above and 5'-RLM-RACE was performed following the Invitrogen GeneRacer protocol with a homemade RNA-linker (Weinberg et al. 2009). First-strand synthesis was performed with a gene-specific primer complementary to the erythromycin resistance cassette coding region (primer 1401, Table 3.1). The resulting cDNA was used as template for PCR with a forward primer specific to the RNA-linker and a nested reverse primer specific to the erythromycin resistance cassette coding region (primers 9 and 1400, Table 3.1). PCR products were cloned into pCR2.1 TOPO-TA vector (Invitrogen) and sequenced to identify transcription start sites.

3.4.6 Growth assays

B. subtilis 168 strains were grown from single colonies in 0.5 mL 2XYT (+ 0.5 μg/mL erythromycin for recombinant strains) in sterile non-treated 24-well cell culture plates for 16-18 hours at 37ºC with shaking (225 rpm). These cultures were used to inoculate 0.5 mL 2XYT (+ 0.5 μg/mL erythromycin for recombinant strains) cultures to a starting OD_{600} ~0.05 in sterile non-treated 24-well cell culture plates. Plates were incubated at 37ºC with shaking (225 rpm) for approximately 24 hours. OD_{600} values were recorded at time points indicated using a SpectraMax M3 Multi-Mode Microplate Reader (Molecular Devices).

For the growth assays at 15ºC*, B. subtilis* 168 strains were grown from single colonies in 2 mL 2XYT (+ 0.5 μg/mL erythromycin for recombinant strains) for 16-18 hours at 37ºC with shaking (225 rpm). Because cultures were not viable in 24-well plates for an extended period of time at 15ºC, these cultures were used to inoculate 25 mL 2XYT (+ 0.5 μg/mL erythromycin for recombinant strains) cultures to a starting OD_{600} ~0.05 in sterile 250 mL flasks. Flasks were incubated at 15ºC with shaking (225 rpm) for approximately 100 hours. OD_{600} values were recorded at time points indicated using a NanoDrop 2000c (Thermo Fisher Scientific).

Doubling times were calculated as previously described using exponential phase $\log_{10}(OD_{600})$ values (Rubinow 1975). The values reported represent the mean of three or more independent replicates; the error reported is the standard error of the mean across biological replicates. To determine the significance, mutant recombinant strain doubling times were compared to those of the wild-type recombinant strain at the corresponding temperature using a Welch's single-tailed T-test in Microsoft Excel. Values were considered significantly different if $p<0.05$. Representative growth curves are shown.

3.4.7 Quantitative RT-PCR

Total RNA was extracted from early-to-mid log (OD_{600} ~0.3-0.7) and early stationary phase (OD_{600} ~2.0-3.0) *B. subtilis* 168 cultures grown in 20 mL 2XYT (+ 0.5 μg/mL erythromycin for recombinant strains) at both 37ºC and 15ºC with shaking (225 rpm). Genomic DNA was removed from 5 μg of total RNA by digestion with RQ1 DNase (Promega) at 37ºC for 40 minutes, followed by heat inactivation at 98ºC for 2 minutes, phenol-chloroform extraction, and ethanol precipitation. Reverse transcription was performed using the DNase-treated RNA, random hexamer, and SuperScript III according to the manufacturer's protocol (Invitrogen). qPCR was conducted with the resulting cDNA using an ABI 7500 Fast Real-Time PCR system and SYBR green detection (ThermoFisher Scientific). *infC* operon transcript expression was quantified using primers targeting the *rplT* coding region and expression of *nifU* was used as an internal normalization control (Table 3.1) (Reiter et al. 2011). Experiments were repeated using reactions lacking reverse transcriptase to confirm removal of genomic DNA. Error bars represent the standard error of the mean across three technical replicates propagated using previously described calculations (Taylor 1997).

3.4.8 Primer extension assays

Log phase total RNA was extracted from *B. subtilis* 168 cultures grown at both 37ºC and 15ºC as described above. For sequencing reaction templates, regions corresponding to the unprocessed 5' ends of both the 16S and 23S rRNAs from the *rrnW* operon were PCR-amplified from *B. subtilis* 168 genomic DNA using forward primers that included the T7 promoter sequence (Table 3.1). PCR products were gel-purified, cloned into pCR2.1 TOPO-TA vector (Invitrogen) for sequencing, and RNA was transcribed from these constructs using T7 RNA polymerase and purified by 6%

denaturing PAGE (Milligan et al. 1987). Synthetic oligonucleotide DNA primers (20 pmol, IDT) complementary to the mature 5' ends of the 16S and 23S rRNAs were 5'-end labeled with [γ-^{32}P] ATP (Perkin Elmer) and purified via 12% denaturing PAGE (Regulski and Breaker 2008).

For the primer extension reactions, 5 μg of total RNA or 1 pmol of *in vitro* transcribed RNA was combined with 1 μl of ^{32}P-labeled primer (~30,000-50,000 cpm/μl) for a final volume of 12 μl in water. This mixture was denatured at 75ºC for 4 minutes and then flash frozen in a dry ice/ethanol bath for 2 minutes before being transferred to ice. SuperScript III (1 μl or 200 U, Invitrogen) and 7 μl of the following master mix were then added to each reaction, for a total reaction volume of 20 μl and a final reaction concentration of: 50 mM Tris HCl [pH 8.3], 75 mM KCl, 3 mM $MgCl_2$, 5 mM DTT, 0.5 mM each dNTP, and 20 U SUPERase-In (Invitrogen). For the sequencing reactions, 1 μl of 100 mM ddNTP (TriLink Biotechnologies) was also added to the appropriate reaction. Reactions were incubated at 55ºC for 30 minutes, stopped with 20 μl of formamide loading dye (95% formamide, 20 mM EDTA [pH 8.0], 0.05% bromophenol blue, 0.05% xylene cyanol), and 8 μl of each sample was separated by 10% denaturing PAGE. Prior to loading, samples were heated at 75ºC for 4 minutes and then cooled on ice for 2 minutes. Gels were dried, exposed to phosphor screens for 48-72 hours, and imaged using a Typhoon FLA 9500 scanner (GE Life Sciences) (modified from (Britton et al. 2007)). Primer extension reactions were performed on RNA extracted from three or more independent biological replicates of each strain, grown to log phase at both 37ºC and 15ºC. Representative gels are shown.

3.4.9 Ribosome sedimentation profiles

B. subtilis 168 strains were grown from single colonies in 2 mL 2XYT (+ 0.5 μg/mL erythromycin for recombinant strains) for approximately 16-18 hours at 37ºC with shaking (225 rpm). These cultures were used to inoculate 250 mL 2XYT (+ 0.5 μg/mL erythromycin for recombinant strains) cultures to a starting OD_{600} ~0.05 in sterile 1L flasks. Flasks were incubated at 15ºC with shaking (225 rpm) until an OD_{600} ~0.3-0.4 was reached. Samples were prepared as previously described (Li et al. 2014). Briefly, cells were harvested by rapid filtration using BioTrace NT pure nitrocellulose 0.2 μm membrane filters (Pall Life Sciences) and flash frozen in liquid nitrogen. Cell pellets were combined with 550 μl of frozen droplets of lysis buffer (20 mM Tris [pH 8.0], 100 mM NH_4Cl, 10 mM $MgCl_2$, 0.4% Triton X-100, 0.1% NP-40, 1 mM chloramphenicol, 100 U/mL DNase I) in 10 mL canisters (Retsch) pre-chilled in liquid nitrogen and pulverized by mixer milling (Qiagen Tissuelyzer II). The resulting lysates were thawed and clarified by centrifugation, and 200 μl of the supernatants were layered onto 10-55% (w/v) sucrose gradients prepared in 20 mM Tris [pH 8.0], 100 mM NH_4Cl, 10 mM $MgCl_2$, and 1 mM chloramphenicol. Gradients were centrifuged at 35000 rpm at 4ºC for 2.5 hours (Beckman Coulter) and then analyzed at 254 nm using a Biocomp Gradient Station iP, Bio-Rad Econo UV Monitor, and Gradient Profiler software.

Table 3.1. Oligonucleotides used in this study.
For each primer pair, the forward primer is listed first and the reverse primer is listed second.

Name	Sequence (5'-3')	Notes
1349	GGCCC**GAATTC**TATGGAGGTGGCTCATGATTAGCAAAGATC	Primer for cloning *infC-rpmI-rplT* operon into pYH213 protein overexpression vector (Yakhnin et al. 2015); EcoRI restriction site in bold; ATT to ATG is underlined
1351	GCCGG**CTGCAG**TTACTTGTTTAATTGAG	Primer for cloning *infC-rpmI-rplT* operon into pYH213 protein overexpression vector (Yakhnin et al. 2015); PstI restriction site in bold
1357	GTTGATCAGTCAACTTATCTGTATAG	Primer for PCR checks and sequencing of pYH213 protein overexpression constructs
1358	CGTTTAAGGGCACCAATAACTG	Primer for PCR checks and sequencing of pYH213 protein overexpression constructs
866	CGCGC**GAATTC**TTGACTAAAGATCCGGTATTGTGTAG	Primer for cloning L20-interacting RNA into modified pDG1728 reporter vector (Babina et al. 2017); EcoRI restriction site in bold
867	GCGC**GGATCC**ATTAACCAATTGATCTTTGCTAATCAT	Primer for cloning L20-interacting RNA into modified pDG1728 reporter vector (Babina et al. 2017); BamHI restriction site in bold; ATT to ATG is underlined
204	TATCTCTTGCCAGTCACGTTACG	Primer for PCR checks and sequencing of pDG1728 reporter constructs
122	GGGGACGACGACAGTATCGGCCTC	Primer for PCR checks and sequencing of pDG1728 reporter constructs
691	CGA**GGATCC**TCGCTTTCTGCTCTTTTTGGATTG	Primer for amplifying 5'-500 bp region of L20-interacting RNA recombinant construct; BamHI restriction site in bold, if needed
692	CTTTAGGGTTATCGAATTCGATAAGCTTCTAGGACGACTTATCCGGAACAACTTTTACATTG	Primer for amplifying 5'-500 bp region of L20-interacting RNA recombinant construct
693	CCCTAGCGCCTACGGGGAATTTGTATCGCGGCCGCTTGACTAAAGATCCGGTATTGTGTAG	Primer for amplifying 3'-500 bp region of L20-interacting RNA recombinant construct
774	GCTCAAATCGGAACTTACCGTAG	Primer for amplifying 3'-500 bp region of L20-interacting RNA recombinant construct

681	TAGAAGCTTATCGAATTCGATAACCCTAAAG	Primer for amplifying erythromycin resistance cassette from pDG1663 (Guérout-Fleury et al. 1996)
682	GCGGCCGCGATACAAATTAAAAGTAGGCG	Primer for amplifying erythromycin resistance cassette from pDG1663 (Guérout-Fleury et al. 1996)
720	CGTCATCAGGAAGCTCTGAAGCTG	Primer for confirming genomic integration of recombinant constructs
775	GAATTTAATCGCATTGCGCAATTTCGTG	Primer for confirming genomic integration of recombinant constructs
745	GCAATGAAACACGCCAAAGTAAAC	Primer for PCR checks, sequencing recombinant L20-interacting RNA constructs
1081	GAATTGAATATAAATCACTTGCAGAAGCACCCGCTTC	Mutagenesis primer for M1, protein-binding mutation
1082	GAAGCGGGTGCTTCTGCAAGTGATTTATATTCAATTC	Mutagenesis primer for M1, protein-binding mutation
746	GAAGCACCCGCTTCTCAGGTGATTGACACATGC	Mutagenesis primer for M2, protein-binding mutation
747	GCATGTGTCAATCACCTGAGAAGCGGGTGCTTC	Mutagenesis primer for M2, protein-binding mutation
1294	GACCGTACATTTTTACCGATACAGATGTTCGTAG	Mutagenesis primer for M3, control mutation
1295	CTACGAACATCTGTATCGGTAAAAATGTACGGTC	Mutagenesis primer for M3, control mutation
1412	GGGTGTTTTATAATGCGGTGCATTTTTGTTTGCCTGC	Mutagenesis primer for M4, terminator mutation
1413	GCAGGCAAACAAAAATGCACCGCATTATAAAACACCC	Mutagenesis primer for M4, terminator mutation
1692	GGAAACTATGCTTTCCGTGACC	Primer for qRT-PCR targeting the *rplT* coding region
1693	TAAGAAAGGCCGTTCATGCG	Primer for qRT-PCR targeting the *rplT* coding region
1546	TTTTACTTCGTGACGGCGGT	Primer for qRT-PCR targeting *nifU*, the normalization control
1547	TTGTTGAACTTGGGCAGCTG	Primer for qRT-PCR targeting *nifU*, the normalization control
8	TAATACGACTCACTATAGG	T7 promoter primer for generating 5'-RACE RNA-linker
7	TTTCTACTCCTTCAGTCCATGTCAGTGTCCTCGTGCTCCAGTCGCCTATAGTGAGTCGTATTA	Primer for generating 5'-RACE RNA-linker
1401	CAGATAGATGTCAGACGCATGGC	Erythromycin resistance cassette outer reverse primer for cDNA synthesis (RT)

9	GACTGGAGCACGAGGACACTGA	5’-RACE RNA-linker forward primer for PCR
1400	GCCAGTTTCGTCGTTAAATGCCC	Erythromycin resistance cassette nested reverse primer for PCR
1107	GCGGGTGCTTCTGCTTGTGATTTATAT	Northern blot probe for L20-interacting RNA (Choonee et al. 2007); hybridize and wash at 42ºC
1402	GTTTACTTTGGCGTGTTTCATTGC	Northern blot probe for erythromycin resistance cassette coding sequence; hybridize and wash at 37ºC
1448	CTTTAGGGTTATCGAATTCGATAAGCTTCTAGAGAGAAAAAGAAAATCTTTCATCCCCAC	Primer for amplifying 5'-500 bp region of L20-interacting RNA re-designed recombinant construct; use with 691
1449	CGCCTACGGGGAATTTGTATCGTTTAAACTTGCGCTCATAGAAAACCCATGTTACAATG	Primer for amplifying 3'-500 bp region of L20-interacting RNA re-designed recombinant construct; use with 774
1444	CTGACAGCTTCCAAGGAGCTAAAGAGGTCTCCTGTTGATAGATCCAGTAATGACC	Primer for amplifying double terminator construct from pYH213 (Yakhnin et al. 2015) for appending onto 3’ end of erythromycin resistance cassette from pDG1663 for re-designed recombinant constructs
1445	GGTCATTACTGGATCTATCAACAGGAGACCTCTTTAGCTCCTTGGAAGCTGTCAG	Primer for amplifying double terminator construct from pYH213 (Yakhnin et al. 2015) for appending onto 3’ end of erythromycin resistance cassette from pDG1663 for re-designed recombinant constructs
1446	GTTTAAACGATACAAATTCCCCGTAGGCGCTAGGGAAAAAAATTACGCCCCGCCCTGCC	Primer for appending double terminator construct onto 3’ end of erythromycin resistance cassette from pDG1663 for re-designed recombinant constructs; use with primer 681
1109	CCAAGTAATACGACTCACTATAGGAATTGAATATAAATCACAAGCAGAAG	T7 *in vitro* transcription primer for *infC* operon transcription start; T7 promoter sequence is underlined
1110	GAATGCATTTTGCAGGC	T7 *in vitro* transcription primer for terminated *infC* operon transcript
1296	GTTGCCTCATCCTTTATATAG	T7 *in vitro* transcription primer for full-length *infC* operon transcript

1621	CCAAGTAATACGACTCACTATAGGCTTTAACAAAGCGGACAAACAAAATGATC	T7 *in vitro* transcription primer for 16S rRNA sequencing reaction template (modified from (Britton et al. 2007)); T7 promoter sequence is underlined
1432	CAGCGTTCGTCCTGAGCCAG	Primer for 16S rRNA sequencing reaction template, primer extension reactions (modified from (Britton et al. 2007))
1623	CCAAGTAATACGACTCACTATAGGACCTTGGGTCTTATAAACAGAACG	T7 *in vitro* transcription primer for 23S rRNA sequencing reaction template (modified from (Redko et al. 2008)); T7 promoter sequence is underlined
1434	CATCGGCTCCTAGTGCCAAGGCATC	Primer for 23S rRNA sequencing reaction template, primer extension reactions (modified from (Redko et al. 2008))

Chapter 4

In vivo behavior of the tandem glycine riboswitch in *Bacillus subtilis*

The content in this Chapter is adapted from the following publication:

Babina AM, Lea NE, Meyer MM. 2017. *In vivo* behavior of the tandem glycine riboswitch in *Bacillus subtilis*. *mBio* 8: e01602-17.

Author contributions:

AMB designed and performed the experiments, analyzed data, and wrote the manuscript. NEL generated select reporter and mutant recombinant strains, optimized select experimental protocols, and collected preliminary data. MMM conceived of the project, designed experiments, assisted with the growth curves, analyzed data, and wrote the manuscript.

4.1 Introduction

The glycine riboswitch regulates the expression of glycine metabolic and transport genes in bacteria. It was among the first riboswitches discovered, and since then, more than 7000 homologs have been identified across numerous bacterial species (Mandal et al. 2004; Ruff et al. 2016). Many examples of the glycine riboswitch, including those found in *B. subtilis* and *Vibrio cholerae*, are composed of two tandem glycine-binding aptamers followed by a single expression platform. Currently, the glutamine riboswitch is the only other known tandem riboswitch in which two or more separate homologous metabolite-sensing aptamers are proposed to act together on a single expression platform (Ames and Breaker 2011). Other examples of tandem riboswitches typically consist of two or more complete and functionally independent riboswitches (Welz and Breaker 2007; Poiata et al. 2009; Zhou et al. 2016).

Due to its unique architecture, the structure and molecular dynamics of the tandem glycine riboswitch have been the subject of numerous biochemical and biophysical studies (Mandal et al. 2004; Lipfert et al. 2007; Kwon and Strobel 2008; Lipfert et al. 2010; Huang et al. 2010; Erion and Strobel 2011; Butler et al. 2011; Kladwang et al. 2011; Sherman et al. 2012; Baird and Ferré-D'Amaré 2013; Esquiaqui et al. 2014; Ruff and Strobel 2014; Ruff et al. 2016). Initial experiments demonstrated cooperative glycine-binding behavior between the two homologous aptamers *in vitro* (Mandal et al. 2004; Kwon and Strobel 2008; Erion and Strobel 2011), and subsequent work proposed a model of sequential glycine binding and asymmetrical cooperativity (Lipfert et al. 2007; Kwon and Strobel 2008; Lipfert et al. 2010; Huang et al. 2010; Erion and Strobel 2011; Butler et al. 2011). Later studies identified a highly conserved leader-linker kink-turn interaction that promotes riboswitch folding and glycine binding (Kladwang et al. 2011; Sherman et al. 2012; Baird and Ferré-D'Amaré 2013; Esquiaqui

et al. 2014). These results suggested that the full-length tandem glycine riboswitch did not demonstrate cooperative binding and the observed cooperativity was an artifact of the truncated constructs utilized for *in vitro* characterization. The current model for tandem glycine riboswitch function is based on extensive analysis of the glycine-binding and dimerization affinities of the two aptamers and proposes that aptamer dimerization and ligand binding are linked equilibria; dimerization interactions promote glycine binding and subsequent glycine binding further stabilizes riboswitch tertiary structure to allow for gene control (Ruff and Strobel 2014). Furthermore, recent work with naturally occurring "singlet" glycine riboswitches (one aptamer followed by a single expression platform) demonstrated that singlet riboswitches bind glycine with affinities comparable to those with the tandem aptamer architecture. However, singlet glycine riboswitches still require interactions between the aptamer domain and a flanking stem-loop "ghost aptamer" for proper folding and ligand-binding activity (Ruff et al. 2016).

While the *in vitro* techniques applied to the tandem glycine riboswitch provide invaluable insight into structure and mechanism of action, such experiments do not always accurately reflect behavior within the cell. Therefore, to examine the tandem glycine riboswitch in a more biologically relevant context, we characterized the expression changes resulting from a panel of *B. subtilis* glycine riboswitch mutants designed to probe aspects of the *in vitro* folding models using β-galactosidase reporter assays. To understand the impact such changes have on organism fitness, we introduced these mutations into the native locus of the tandem glycine riboswitch preceding the *gcvT* glycine cleavage operon within the *B. subtilis* genome and examined organism phenotype under a variety of conditions.

Our data suggest that mutations disrupting first aptamer tertiary structure have the greatest impact on tandem glycine riboswitch regulation and gene expression. We

find that glycine-induced expression of the *gcvT* operon is necessary for *B. subtilis* growth, swarming motility, and biofilm formation in high glycine environments. However, constitutive expression of the *gcvT* operon in the absence of glycine also has adverse effects on growth, emphasizing the importance of the tandem glycine riboswitch function as a genetic “on” switch in response to excess glycine.

4.2 Results

4.2.1 Double ligand occupancy is not necessary to elicit a regulatory response

To investigate the impacts of disruptions to glycine-binding and inter-aptamer interactions on glycine riboswitch *in vivo* regulation, we examined a number of mutations to the tandem riboswitch structure using β-galactosidase reporter assays in *B. subtilis* (Figure 4.1A). Each mutant riboswitch sequence, including the first nine codons of *gcvT*, was cloned in-frame as a translational fusion with a *lacZ* reporter under the control of the native *gcvT* operon promoter, and then stably integrated as a single copy into the *amyE* locus of the *B. subtilis* 168 genome. The regulatory properties of each riboswitch mutant were then assessed by measuring β-galactosidase activity of reporter strains grown in varying glycine concentrations. As expected, the wild-type riboswitch construct behaved as a genetic “on” switch in the presence of glycine (Figure 4.1B). The β-galactosidase activity of the wild-type riboswitch reporter strain increased approximately 17-fold upon the addition of 0.25% glycine to the medium, and this activity level was maintained for all subsequent glycine concentrations tested.

Mutations M1 and M2 target the glycine-binding domains on the first and second aptamer, respectively (Figure 4.1A). These bases have been shown to be important for glycine binding via inline probing experiments with both the *B. subtilis* and *V. cholerae* riboswitches and are in close proximity to the ligand in the crystal structure of the

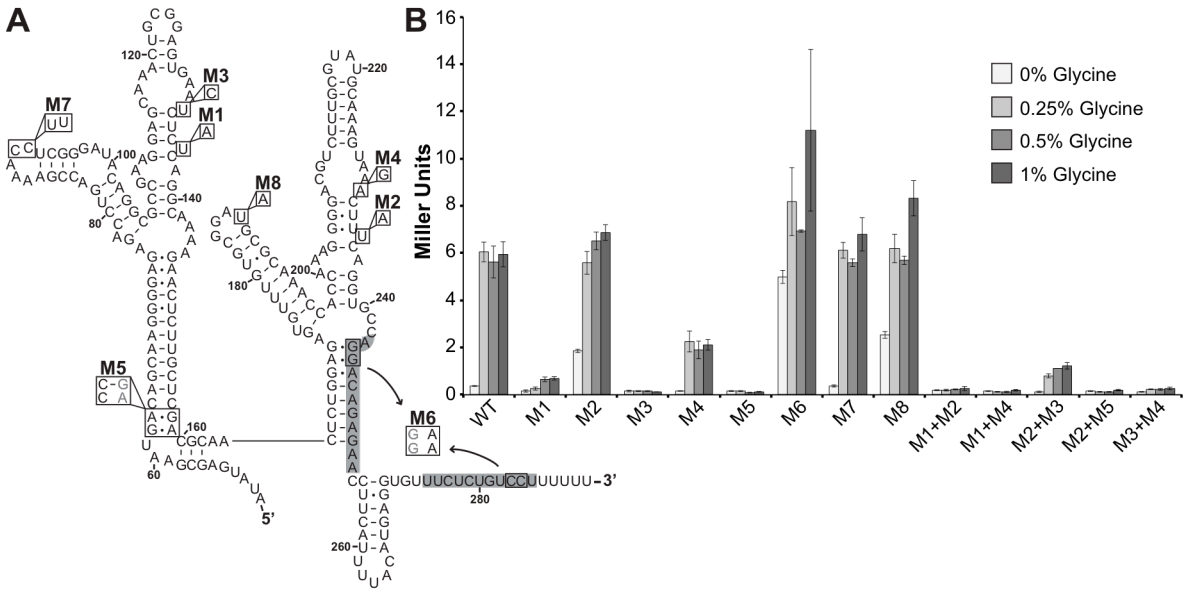


Figure 4.1. Regulatory activity of glycine riboswitch mutations. (A) Secondary structure of the *B. subtilis* glycine riboswitch with mutations M1-M8. Nucleotides are numbered from the transcript start site, +1 (Irnov et al. 2010). Gray shading highlights nucleotides that base pair to form the transcription terminator stem when the riboswitch is in the "off" conformation. (B) β-galactosidase activity (in Miller Units) of riboswitch mutant constructs in the presence of increasing glycine concentrations. The values reported represent the mean of three or more independent experimental replicates; error bars represent standard error of the mean across biological replicates. WT, wild-type reporter.

Fusobacterium nucleatum RNA (Mandal et al. 2004; Kwon and Strobel 2008; Erion and Strobel 2011; Butler et al. 2011). Additionally, previous *in vitro* work with homologous *V. cholerae* glycine riboswitch mutants demonstrated that each binding-site mutation disrupts glycine binding independently and that the ability of one aptamer to bind glycine only has a small effect on the glycine-binding affinity of the other aptamer (Ruff and Strobel 2014).

We find that mutation M1 significantly reduced, but did not completely abolish riboswitch responsiveness to glycine and downstream gene expression. The maximum Miller Units measured with the M1 construct were approximately one-tenth of those obtained with the wild-type construct (0.7 ± 0.06 Miller Units versus 5.9 ± 0.52 Miller Units, respectively, in 1% glycine). Surprisingly, the M2 mutation to the glycine-binding pocket on the second aptamer did not affect riboswitch regulation in response to glycine

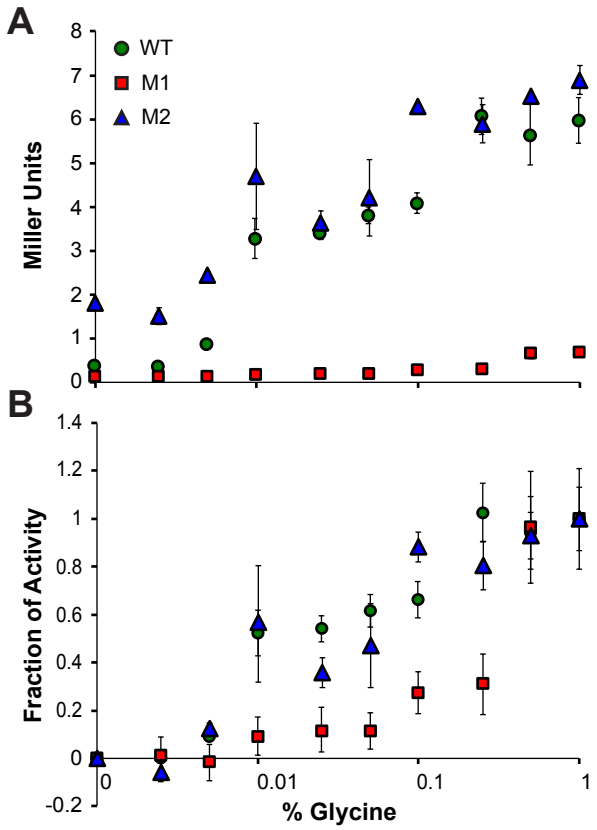


Figure 4.2. Sensitivity and dynamic range of the M1 and M2 glycine riboswitch reporter constructs. (A) β-galactosidase activity (in Miller Units) of the glycine-binding mutant riboswitch reporter constructs in the presence of increasing glycine concentrations. (B) Data in A normalized to represent fraction of β-galactosidase activity for the glycine riboswitch reporter constructs at each glycine concentration tested. The values reported represent the mean of two or more independent experimental replicates. Error bars represent standard error of the mean across biological replicates; some error bars may be smaller than the data points. For both A and B, the x-axis is graphed on a $\log_{10}$ scale. WT, wild-type reporter.

or maximum gene expression levels (6.9 ± 0.33 Miller Units compared to 5.9 ± 0.52 Miller Units from wild type in 1% glycine) (Figure 4.1B).

This is in contrast to previous *in vitro* studies with the *V. cholerae* riboswitch that demonstrate disruptions to the glycine-binding site on the second aptamer have the greatest impact on the binding affinity of the first aptamer (Ruff and Strobel 2014). The M2 single mutation also resulted in higher basal β-galactosidase activity in comparison to the wild-type construct at 0% glycine. The proximity of the M2 mutation to the adjacent terminator stem may affect terminator stability, resulting in the elevated basal constitutive expression. In agreement with published findings, the M1+M2 double glycine-binding mutant construct abrogated regulation (Ruff and Strobel 2014).

Finally, to determine if mutations to the glycine-binding domain on one aptamer influence the glycine-binding sensitivity and dynamic range of the other aptamer, we

measured β-galactosidase activity from the wild-type, M1, and M2 glycine riboswitch reporter strains grown in range of additional glycine concentrations (Figure 4.2). Apart from the high basal β-galactosidase activity characteristic of the M2 mutation at low glycine concentrations, the M2 mutant construct behaved similarly to the wild-type riboswitch under all conditions tested. This indicates that the glycine-binding sensitivity of the first aptamer is comparable to that of a tandem glycine riboswitch construct in which both glycine-binding domains remain intact. Basal β-galactosidase activity was recorded with the M1 construct for the majority of the glycine concentrations tested during this experiment. The modest regulatory response previously measured for the M1 reporter was observed only for glycine concentrations greater than 0.1%. Thus, the glycine-binding ability of the second aptamer does not appear to have the same range of sensitivity as the first aptamer. Altogether, based on the behavior of the M1 and M2 single and double mutant constructs, glycine must bind to at least one aptamer to promote a regulatory response. However, glycine binding to the first aptamer is necessary to drive maximum downstream gene expression and allows for a more sensitive regulatory response to subtle changes in glycine concentration.

4.2.2 Inter-aptamer interactions are important for tandem glycine riboswitch regulation

The M3 and M4 mutations were designed based on the structural data of the homologous *F. nucleatum* riboswitch and target the U-A γ inter-aptamer contacts that are proposed to play a crucial role in communicating the status of glycine binding between the two aptamers (Figure 4.1A) (Butler et al. 2011). Mutating this dimerization interface on the first aptamer (M3) resulted in a complete loss of regulation. However, the M4 mutation to the second aptamer retained regulatory activity in response to glycine,

although the maximum Miller Units obtained were about one-third of those measured with the wild-type riboswitch construct (Figure 4.1B).

This finding is the reverse of what has been reported with previous *in vitro* mutational analyses of the dimerization interface. A point mutation homologous to M4 on the *V. cholerae* riboswitch was found to have the most severe impact on aptamer dimerization via *trans* gel-shift assays (Ruff and Strobel 2014). Mutating the γ dimerization interface on both aptamers (M3+M4) abrogated glycine riboswitch function and yielded basal expression levels that were slightly higher than those measured for the M3 single mutant construct (0.27 ± 0.06 Miller Units for the M3+M4 construct compared to 0.12 ± 0.02 Miller Units for the M3 construct at 1% glycine).

To examine the importance of inter-aptamer interactions for glycine binding and consequent regulation, we combined each dimerization mutant with the mutation to the glycine-binding pocket on the opposite aptamer (M1+M4 and M2+M3). The M1+M4 mutant displayed total loss of regulation and downstream expression, whereas the M2+M3 double mutant retained regulatory activity and modest downstream expression. Similar to the single dimerization mutants (M3, M4), the behavior of the M1+M4 and M2+M3 double mutant constructs was the opposite of what has been previously observed. Equilibrium dialysis assays with a *V. cholerae* glycine riboswitch construct homologous to our M2+M3 construct resulted in the greatest reduction in glycine-binding affinity (Ruff and Strobel 2014). Taken together, we find that although regulation can occur with a disrupted dimerization interface (M4) and in combination with loss of glycine-binding to the second aptamer (M2+M3), proper tandem glycine riboswitch function and maximum gene expression appear to depend on the stabilization provided by inter-aptamer interactions and glycine binding to the first aptamer.

4.2.3 The leader-linker kink-turn is required for tandem glycine riboswitch function

To investigate the impact of the leader-linker interaction, the M5 mutation disrupts the kink-turn that forms the P0 helix found in over 90% of tandem glycine riboswitches (Figure 4.1A) (Kladwang et al. 2011; Sherman et al. 2012). Previous inline probing, native gel analysis, small-angle X-ray scattering, and isothermal titration calorimetry experiments showed that this leader-linker interaction is formed independent of glycine and results in a more stable and compact structure that enhances glycine-binding and inter-aptamer interactions (Kladwang et al. 2011; Sherman et al. 2012; Baird and Ferré-D'Amaré 2013; Esquiaqui et al. 2014). Truncations of the 5' leader and mutations to the linker region that disrupt the formation of the P0 helix in the *V. cholerae* riboswitch significantly reduced the ligand-binding affinity of the RNA *in vitro*. The M5 mutation disrupting the kink-turn resulted in a complete loss of regulation. Combining the M5 mutation with the glycine-binding mutation on the second aptamer (M2+M5) also resulted in loss of regulation and yielded Miller Units similar to that of the M5 single mutant construct and the M1+M4 double mutant (Figure 4.1B). This supports previous findings indicating that the leader-linker kink-turn plays a key role in riboswitch-mediated regulation and is important for glycine binding to the first aptamer.

4.2.4 Control mutations to the glycine riboswitch behave as anticipated

The M6 mutation destabilizes the base of the Rho-independent terminator stem that forms when the riboswitch is in the "off" conformation in the absence of glycine. As expected, this mutation allowed constitutive expression of the *lacZ* reporter (Figure 4.1). The slight increase in β-galactosidase activity in the presence of glycine (~2-fold

increase) can be attributed to further destabilization of the terminator upon glycine binding and aptamer dimerization, as these domains remain intact.

We designed mutations M7 and M8 as controls that target regions on the first and second aptamer, respectively, in which RNA structure is conserved, but nucleotide composition varies (Figure 4.1) (Mandal et al. 2004). Riboswitch regulation remained intact for both of these mutant constructs with Miller Units comparable to that of the wild-type riboswitch under all glycine concentrations tested. The basal β-galactosidase activity of the M8 mutant construct in the absence of glycine was modestly higher than that of the wild-type and M7 mutant riboswitch constructs. Like the M2 single mutation, this mutation may also result in a slight change in the conformation of the second aptamer, possibly affecting its glycine-binding affinity, dimerization interactions, or the stability of the adjacent terminator stem when in the "off" conformation.

4.2.5 *gcvT* expression of native locus recombinant strains reflects β-galactosidase assay data

In *B. subtilis*, the tandem glycine riboswitch turns on expression of the *gcvT* operon (encoding components of the glycine cleavage system, *gcvT*, *gcvPA*, and *gcvPB*, that catabolize glycine to ammonia, carbon dioxide, and one carbon units utilized via the folate pool) in response to glycine (Mandal et al. 2004; Kikuchi et al. 2008; Tezuka and Ohnishi 2014). To explore the physiological role of the glycine riboswitch and determine whether *gcvT* operon expression changes resulting from mutations to the riboswitch impact *B. subtilis* fitness and growth, we replaced the native copy of the glycine riboswitch within the *B. subtilis* NCIB 3610 genome with either a wild-type or mutant recombinant version (Figure 4.3A). A recombinant strain in which the entire *gcvT-gcvPB* locus was deleted was also generated to serve as a negative control (Δ*gcvT-gcvPB*). To

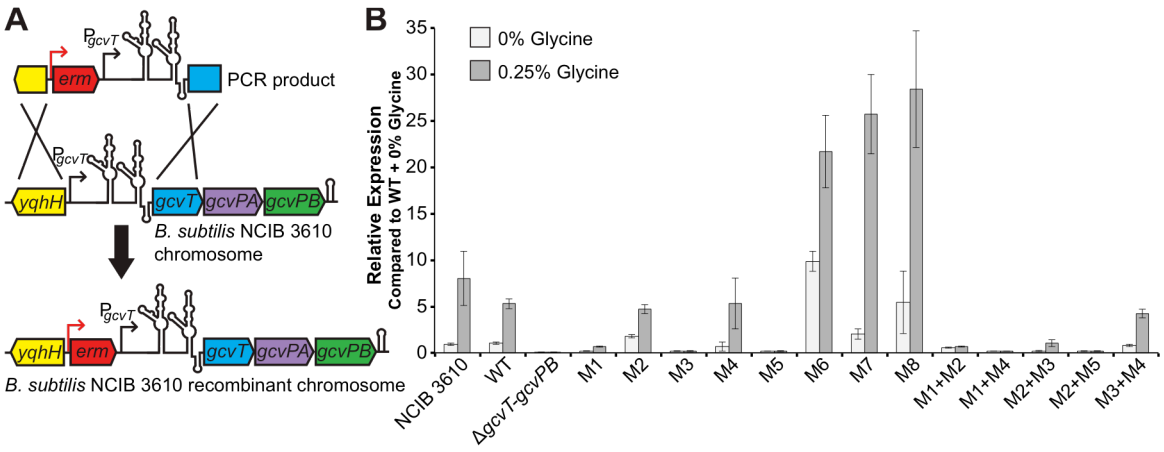


Figure 4.3. Construction and confirmation of recombinant glycine riboswitch *B. subtilis* strains. (A) Schematic of the strategy used to generate the recombinant *B. subtilis* NCIB 3610 strains. The *gcvT* operon promoter (P_{gcvT}), glycine riboswitch, and two ~500 bp regions flanking the promoter and riboswitch locus were PCR-amplified from *B. subtilis* 168 genomic DNA. A PCR product in which an erythromycin resistance cassette (*erm*) was introduced into the intergenic region immediately upstream from the *gcvT* operon promoter was generated. Transformation of cloned PCR products into *B. subtilis* NCIB 3610 replaced the native copy of the glycine riboswitch with either a wild-type or mutant recombinant version via double-crossover homologous recombination. (B) qRT-PCR quantification of the native *gcvT* transcript from each of the recombinant glycine riboswitch strains grown in M9 minimal medium +/- 0.25% glycine. For each strain/condition, *gcvT* expression was normalized to expression of the *nifU* control transcript. Graph depicts relative *gcvT* expression from each strain compared to *gcvT* expression from the wild-type recombinant (WT) grown in the absence of glycine (0% glycine). Error bars represent standard error of the mean across three technical replicates. NCIB 3610, parental strain.

confirm our β-galactosidase assay results accurately represent changes in glycine riboswitch regulation and gene expression in the NCIB 3610 mutant recombinant strains, we performed quantitative RT-PCR (qRT-PCR) using primers within the *gcvT* coding region on log phase total RNA from each of the recombinant strains grown in M9 minimal medium +/- 0.25% glycine (Figure 4.3B).

These results are generally in good agreement with our reporter assay data (Figure 4.1B) and further confirm the behavior of the *B. subtilis* tandem glycine riboswitch as an “on” switch that regulates *gcvT* operon expression via a transcription terminator in response to glycine. The mutations that retained riboswitch function in the β-galactosidase assays (M1, M2, M4, M2+M3) exhibited similar trends in regulatory activity and expression as measured by qRT-PCR. Likewise, mutations M3, M5, M1+M4,

and M2+M5 abrogated glycine riboswitch regulation and resulted in *gcvT* transcript expression levels comparable to that of the Δ*gcvT-gcvPB* operon deletion recombinant strain. The recombinant strain carrying the M6 mutation to the terminator exhibited elevated constitutive expression of the *gcvT* transcript regardless of the presence or absence of glycine. Also consistent with our previous data, the presence of the M1+M2 double mutation removed riboswitch regulation and resulted in basal *gcvT* transcript expression.

The only major discrepancy between the qRT-PCR results and our β-galactosidase assays is the behavior of the M3+M4 mutant. While the β-galactosidase assays suggest the M3+M4 mutation resulted in loss of riboswitch function and slightly elevated basal expression, our qRT-PCR data indicate that riboswitch regulation is retained in the corresponding mutant recombinant strain. Additionally, the control recombinant strains containing mutations M7 and M8 exhibited *gcvT* transcript levels approximately 4 to 5-fold higher than that of the wild-type recombinant strain when grown in 0.25% glycine. This increase in expression is not represented in the reporter assay data where the M7 and M8 mutant constructs yielded Miller Units comparable to that of the wild-type construct in the presence of glycine. These discrepancies may be attributable to differences between *B. subtilis* strains 168 and NCIB 3610, to the increased sensitivity of qRT-PCR relative to β-galactosidase assays, or to the measurement of the native transcript rather than a translational reporter that does not include the entire transcript.

4.2.6 *gcvT* operon expression affects glycine sensitivity and doubling time during planktonic growth

Once the behavior of the glycine riboswitch mutations was confirmed in the recombinant *B. subtilis* NCIB 3610 strains, growth curves were performed with all recombinant strains in M9 minimal medium with increasing glycine concentrations (Table 4.1). High glycine concentrations are known to inhibit bacterial growth, as excess glycine interferes with cell wall biosynthesis (Snell and Guirard 1943; Gordon et al. 1951; Hishinuma et al. 1969, 1971; Hammes et al. 1973; Tezuka and Ohnishi 2014). In agreement with these observations, we observed a reduction in maximum growth for all recombinant strains in the presence of glycine, and the doubling times increased for nearly all strains as the glycine concentration in the medium was increased (Table 4.1). Prolonged lag phase and abnormal or little to no cell growth was also observed for most strains in 1% glycine; this is reflected in the high standard error reported for this glycine concentration (Hishinuma et al. 1969).

Strikingly, the strain harboring the M6 mutation to the terminator stem grew approximately 1.3 times slower than the wild-type recombinant strain in medium lacking glycine. Furthermore, its doubling time decreased upon the addition of 0.25% glycine to the medium (101 min to 88 min), and growth was comparable to that of the wild-type recombinant strain in all glycine-supplemented conditions. As the M6 mutation results in constitutive expression of the *gcvT* operon, unnecessary expression of the glycine cleavage system in the absence of glycine appears to have adverse effects on cell growth, and this defect is rescued in the presence of glycine.

As expected, the Δ*gcvT-gcvPB* recombinant strain grew comparably to the wild-type recombinant in medium lacking glycine and demonstrated a significant increase in doubling time upon the addition of glycine to the medium, growing approximately 1.3

Table 4.1. Doubling times (min) of recombinant glycine riboswitch *B. subtilis* strains grown in increasing glycine concentrations. Strains were grown in M9 minimal medium with increasing glycine concentrations at 37ºC with shaking (225 rpm) for 24 hours. Values reported are the mean of three or more independent experimental replicates; ± indicates the standard error of the mean across biological replicates. Numbers in parentheses denote strain doubling time relative to that of the wild-type recombinant (WT) strain. Asterisks (*) indicate strains that grew significantly worse than the wild-type recombinant strain at the corresponding glycine concentration ($p<0.05$). NCIB 3610, parental strain.

Strain	0% Glycine	0.25% Glycine	0.5% Glycine	1% Glycine
NCIB 3610	70.0 ± 2.6	80.0 ± 2.7	82.8 ± 1.7	108.9 ± 5.8
WT	76.2 ± 2.3	87.9 ± 2.1	90.3 ± 3.9	110.6 ± 4.4
Δ*gcvT-gcvPB*	73.7 ± 2.7 (1.0)	*110.1 ± 4.8 (1.3)	*118.8 ± 5.1 (1.3)	127.7 ± 9.8 (1.2)
M1	81.7 ± 5.2 (1.1)	98.7 ± 7.5 (1.1)	98.7 ± 11.0 (1.1)	116.5 ± 11.6 (1.1)
M2	69.3 ± 3.2 (0.9)	77.9 ± 4.5 (0.9)	84.1 ± 3.8 (0.9)	116.3 ± 12.5 (1.1)
M3	*116.8 ± 7.8 (1.5)	*143.6 ± 4.0 (1.6)	*143.4 ± 0.0 (1.6)	*153.4 ± 5.0 (1.4)
M4	76.7 ± 2.8 (1.0)	92.3 ± 2.5 (1.0)	94.3 ± 3.4 (1.0)	110.6 ± 5.0 (1.0)
M5	*84.6 ± 2.8 (1.1)	*115.3 ± 7.8 (1.3)	*122.8 ± 7.0 (1.4)	123.8 ± 13.8 (1.1)
M6	*101.4 ± 7.4 (1.3)	87.8 ± 1.7 (1.0)	96.1 ± 1.0 (1.1)	106.1 ± 7.7 (1.0)
M7	80.0 ± 5.7 (1.1)	91.9 ± 5.6 (1.1)	91.2 ± 6.3 (1.0)	106.7 ± 4.8 (1.0)
M8	78.2 ± 3.5 (1.0)	90.6 ± 3.5 (1.0)	97.3 ± 3.0 (1.1)	110.6 ± 10.3 (1.0)
M1+M2	80.0 ± 1.9 (1.1)	95.5 ± 4.5 (1.1)	103.1 ± 5.0 (1.1)	*133.5 ± 6.8 (1.2)
M1+M4	76.1 ± 6.1 (1.0)	*107.2 ± 4.9 (1.2)	*106.7 ± 2.3 (1.2)	126.9 ± 7.8 (1.2)
M2+M3	80.8 ± 2.8 (1.1)	89.6 ± 3.1 (1.0)	96.4 ± 3.8 (1.1)	116.1 ± 4.3 (1.1)
M2+M5	78.6 ± 1.8 (1.0)	*100.4 ± 1.9 (1.1)	*112.3 ± 6.7 (1.2)	126.8 ± 9.7 (1.2)
M3+M4	*137.4 ± 6.5 (1.8)	*143.6 ± 4.0 (1.6)	*148.1 ± 2.4 (1.6)	*148.1 ± 2.4 (1.3)

times slower than wild-type recombinant in all glycine concentrations tested. This suggests that the *gcvT* glycine cleavage operon is important for glycine detoxification in *B. subtilis*. The M5, M1+M4, and M2+M5 mutant recombinant strains also displayed a modest increase in doubling time relative to the wild-type recombinant upon the addition of glycine to the medium. The increase in glycine sensitivity of these mutant recombinant strains during planktonic growth can be attributed to extremely low expression of the *gcvT* operon (Figure 4.1B, 4.3B), further supporting the role the *gcvT* operon plays in glycine detoxification.

No increase in glycine sensitivity was observed for the majority of the glycine riboswitch mutant recombinant strains that retained regulatory activity (M1, M2, M4, M7, M8, M2+M3). The doubling times of these strains remained consistent relative to that of

the wild-type recombinant strain at all glycine concentrations (Table 4.1). The recombinant strain carrying the M1+M2 double mutation also grew comparably to the wild-type recombinant under all conditions. The basal expression from the M1+M2 construct appears sufficient for normal planktonic growth in high glycine environments.

Interestingly, the strains carrying the M3 and M3+M4 mutations exhibited the longest doubling times, growing approximately 1.6 times slower than the wild-type recombinant strain in the presence and absence of glycine. This growth defect does not correlate well with gene expression as measured by β-galactosidase activity and qRT-PCR. Each of the recombinant strains was created independently from the parental strain, thus it is unlikely that this defect is due to some other common mutation. It is possible that components of the *gcvT* operon are important for metabolic processes not directly related to glycine and that the M3 and M3+M4 mutations further interfere with translation of the native operon transcript. Therefore, the glycine-independent planktonic growth defects observed for these strains may be due to disruptions to other fundamental biochemical processes important for growth.

While inhibition of cell growth was observed for all strains in increasing glycine concentrations as previously reported, only strains with mutations that completely abolish riboswitch function and result in extremely low *gcvT* operon expression appear to have heightened glycine sensitivity during planktonic growth in minimal medium compared to the wild-type recombinant strain. The majority of these glycine-sensitive recombinant strains carry mutations that disrupt glycine binding to the first aptamer and/or first aptamer tertiary structure.

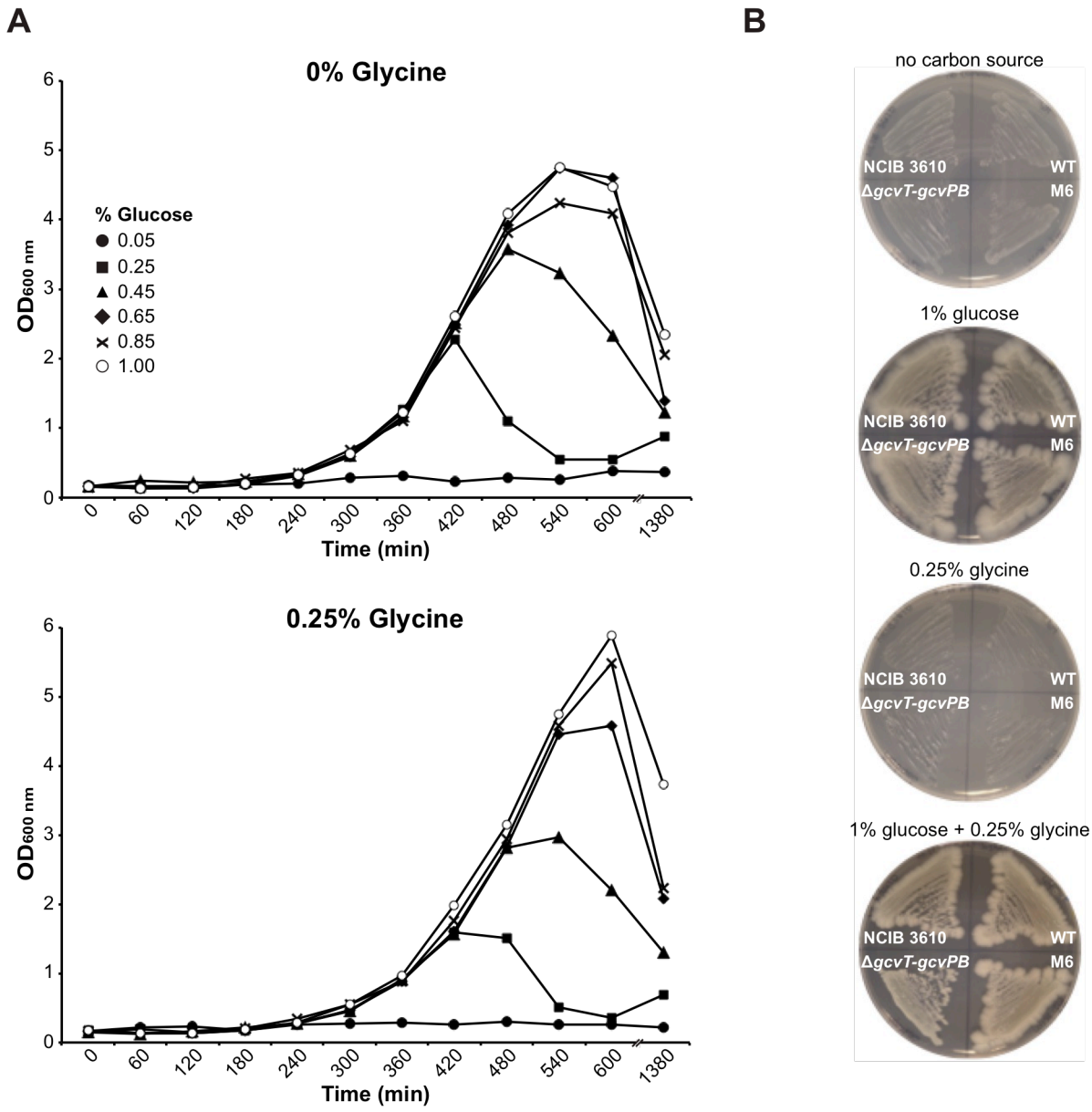


Figure 4.4. *B. subtilis* growth using glucose and glycine as carbon sources. (A) Growth curves of *B. subtilis* NCIB 3610 (parental strain) in M9 minimal medium +/- 0.25% glycine with varying concentrations of glucose. Cultures were incubated at 37ºC with shaking (225 rpm) for approximately 24 hours and OD_{600} measurements were recorded at time points indicated. (B) M9 minimal medium patch plates of select *B. subtilis* NCIB 3610 strains with various carbon sources. Plates were incubated at 37ºC for one week and photographed. Photographs after 48 hours of incubation are shown, however, no significant changes were observed on all plates after one week. WT, wild-type recombinant.

4.2.7 Glycine cannot be utilized as a sole carbon source by *B. subtilis* NCIB 3610

Others have demonstrated that the tandem glycine riboswitch-regulated operons *gcvT-gcvH* and *gcvP* facilitate use of glycine as a carbon source in *Streptomyces griseus* (Tezuka and Ohnishi 2014). To examine whether *B. subtilis* could similarly metabolize glycine and assess whether this phenotype is sensitive to changes in *gcvT* operon expression resulting from glycine riboswitch mutations, we performed growth experiments in both solid and liquid minimal media in the presence and absence of glycine with limiting glucose. We found that in contrast to *S. griseus*, *B. subtilis* NCIB 3610 is not able to grow on glycine as a sole carbon source (Figure 4.4). Of note, *gcvT-gcvH* and *gcvP* exist as separate operons and are under the control of separate tandem glycine riboswitches in *S. griseus*, whereas *gcvT* and *gcvPA-gcvPB* (two genes encoding separate subunits of the glycine decarboxylase GcvP) are co-located in the *B. subtilis* genome and are under the regulation of a single tandem glycine riboswitch. The predicted *B. subtilis* homolog to *gcvH* (also known as *yusH*) occurs elsewhere in the genome and does not appear to be glycine riboswitch controlled.

4.2.8 Glycine riboswitch mutants have reduced swarming motility in high glycine environments

To assess whether more complex *B. subtilis* phenotypes might be more sensitive readouts of glycine toxicity than planktonic growth doubling times, we next investigated the impact of riboswitch mutations and aberrant regulation of the glycine cleavage operon on swarming motility in the presence of glycine. Swarming motility is highly dependent on cellular differentiation into a swarming proficient state and cellular contacts with surfaces and neighboring cells, all of which are at least partially mediated by peptidoglycan remodeling and cell wall synthesis and structural integrity

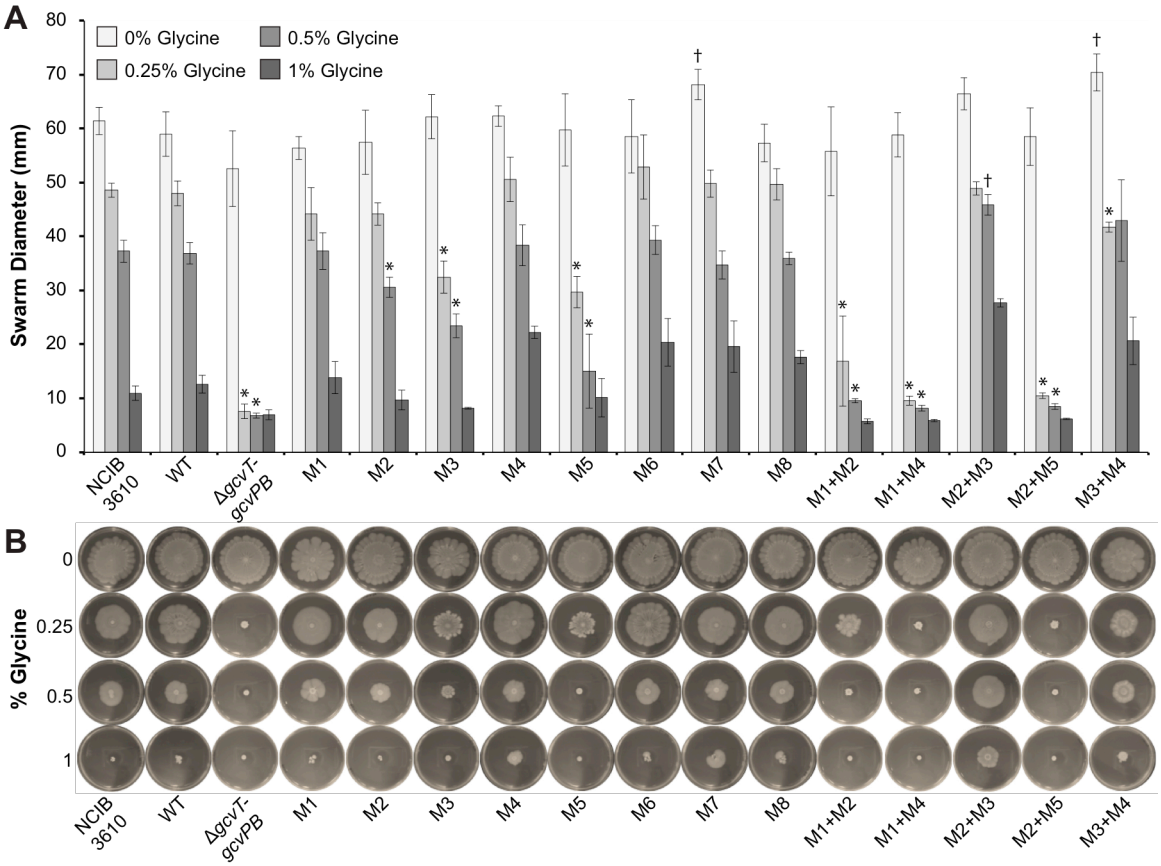


Figure 4.5. Swarming motility of recombinant glycine riboswitch *B. subtilis* strains in increasing glycine concentrations. M9 minimal medium swarm agar plates with varying glycine concentrations were inoculated with each strain and incubated at 37ºC for 48 hours. (A) Swarm diameters of recombinant glycine riboswitch *B. subtilis* strains. Three measurements were taken for each plate and averaged. The values reported represent the mean of three independent experimental replicates; error bars represent the standard error of the mean. Asterisks (*) indicate mutant recombinant strains that grew significantly worse than the wild-type recombinant (WT) in the corresponding glycine concentration. Daggers (†) indicate mutant recombinant strains that grew significantly better than the wild-type recombinant in the corresponding glycine concentration ($p<0.05$). Significance was not determined for measurements recorded at 1% glycine. (B) Representative photographs of swarming motility assays. NCIB 3610, parental strain.

(Kearns and Losick 2003; Copeland and Weibel 2009). Excess glycine can be mis-incorporated into bacterial cell walls and high glycine concentrations inhibit the enzymes responsible for the addition of L-alanine into peptidoglycan precursors, resulting in weakened cell walls and premature lysis (Gordon et al. 1951; Hishinuma et al. 1969, 1971; Hammes et al. 1973).

No significant difference in swarm diameter was observed for all strains in the absence of glycine. Similar to cell growth rate, swarm diameter decreased for all strains as glycine concentration in the medium was increased (Figure 4.5). Little to no migration was observed at 1% glycine for all strains. The large degree of variation in swarm diameter at this concentration can be attributed to the prevalence of escape motile or flare mutants, which resulted in swarms with non-uniform diameters and is indicative of glycine toxicity. Consequently, significance was not determined for swarm diameter measurements at 1% glycine. Apart from differences in migration diameter, no other strain-specific or glycine concentration-dependent trends were noted in swarm morphology.

Strains with mutations that abolish riboswitch regulation and result in low downstream expression (M3, M5, M1+M2, M1+M4, M2+M5) demonstrated a significant reduction in swarming motility in the presence of glycine compared to the NCIB 3610 parental and wild-type recombinant strains (Figure 4.5). The M1+M2, M1+M4, and M2+M5 double mutant recombinant strains exhibited the most severe glycine-sensitive swarm phenotypes that were comparable to those of the Δ*gcvT-gcvPB* recombinant strain. Strains that retain riboswitch function and/or result in elevated constitutive operon expression (M1, M2, M4, M6, M7, M8, M2+M3, M3+M4) all behaved similar to the wild-type recombinant and NCIB 3610 parental strains under all assay conditions.

It is unlikely that the defects observed during planktonic growth play a role in swarm migration distance. While the M3 mutant recombinant strain showed a significant swarm migration defect, the migration distance of the M3+M4 double mutant recombinant strain was comparable to the wild-type recombinant and the NCIB 3610 parental strains at all glycine concentrations. Similarly, while the doubling times of the M1+M2 mutant recombinant strain did not significantly differ from those of the wild-type

recombinant strain during planktonic growth, the M1+M2 recombinant strain demonstrated significant glycine-sensitivity during the swarming motility assays. Defects observed during planktonic cell growth can be distinct from those observed during static growth.

4.2.9 Mutations to the glycine riboswitch inhibit biofilm formation in high glycine environments

To further assess the effects of disruptions to glycine riboswitch function and *gcvT* operon regulation on *B. subtilis* fitness in the presence of excess glycine, we analyzed mutant recombinant strain biofilm formation on solid medium as well as the development of floating biofilms formed at the air-liquid interface (pellicle), as *B. subtilis* is a robust model organism for the study of biofilm development (for review see: (Cairns et al. 2014)). A recent study demonstrated that interfering with cell wall composition greatly disrupts *B. subtilis* biofilm formation (Bucher et al. 2015) and excess glycine has been shown to inhibit *Streptoccocus sobrinus* aggregation by way of reduced glucan-binding ability and weakened cell wall integrity (Luengpailin and Doyle 2000).

Extent of biofilm assembly on both solid and liquid media was reduced for all strains as glycine concentration was increased. Little to no pellicle development was observed for all strains at 1% glycine; the pellicles that did form were very fragile and had a smooth surface (Figure 4.6A,B). Consequently, significance is not reported for the crystal violet staining assays at this concentration. Similarly, biofilms that developed on solid medium with 1% glycine were small in diameter and lacked the wrinkled phenotype characteristic of robust *B. subtilis* biofilms (Figure 4.6C).

The M3 and M5 single mutant recombinant strains as well as the M1+M2, M1+M4, and M2+M5 double mutant recombinant strains exhibited the greatest defects in

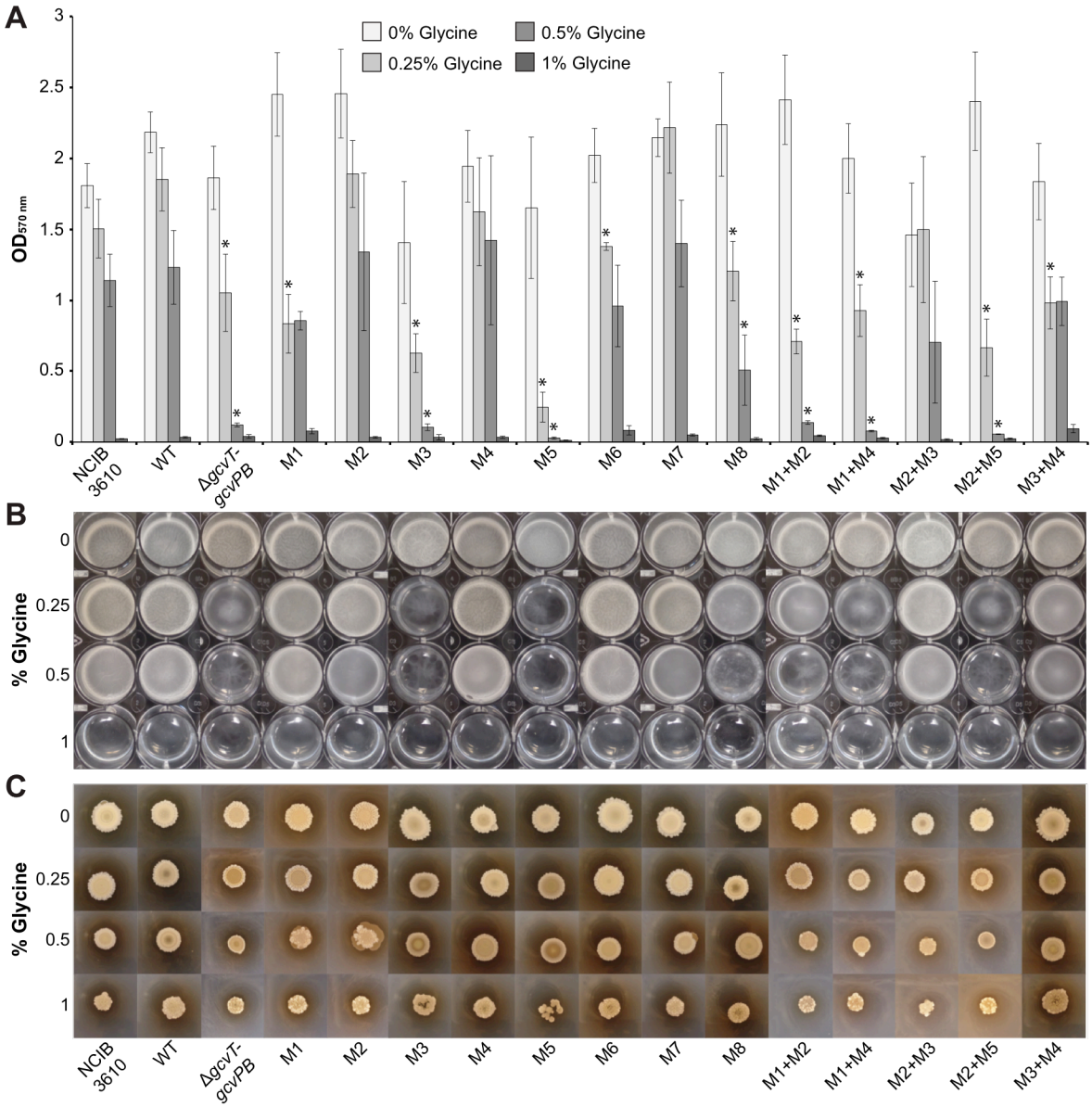


Figure 4.6. Biofilm formation of recombinant glycine riboswitch *B. subtilis* strains in increasing glycine concentrations. (A) Crystal violet staining of recombinant glycine riboswitch *B. subtilis* strain pellicle formation. The values reported represent the mean of six or more independent experimental replicates; error bars represent standard error of the mean across biological replicates. Asterisks (*) indicate mutant recombinant strains that grew significantly worse than the wild-type recombinant (WT) in the corresponding glycine concentration ($p<0.05$). Significance was not determined for measurements recorded at 1% glycine. (B) Representative photographs of recombinant strain pellicle formation after 24 hours incubation at 37ºC. (C) Representative photographs of recombinant strain colony biofilm formation on solid medium after 5 days incubation at 30ºC. NCIB 3610, parental strain.

biofilm formation in the presence of glycine on both solid and liquid media (Figure 4.6). Pellicles were thinner and had significantly reduced or no surface wrinkling in comparison to the other strains. The solid-surface biofilms also displayed defects in colony wrinkling and were often dark brown in color in the presence of glycine, indicating sporulation or cell death due to harsh environmental conditions (Sandman et al. 1988). The biofilms formed by these mutant recombinant strains were comparable to those of the Δ*gcvT-gcvPB* recombinant strain under all conditions.

The recombinant strains in which glycine riboswitch function remains intact and/or the *gcvT* operon is constitutively expressed (M1, M2, M4, M6, M7, M8, M2+M3, M3+M4) generated robust biofilms on both media types with morphologies similar to those produced by the wild-type recombinant and NCIB 3610 parental strains. It should be noted that while crystal violet staining is a simple and high-throughput method for quantifying pellicle biofilm formation, assay results are highly sensitive to minor changes in protocol or conditions and often yield a great range of variation between replicates (Li et al. 2003; Kwasny and Opperman 2010), as demonstrated by the high standard error reported. Nevertheless, these biofilm assay results and observed trends in glycine sensitivity are consistent with those from our swarming motility experiments.

4.3 Discussion

Since the initial discovery of the tandem glycine riboswitch, numerous studies have sought to elucidate the structure, dynamics, and selective advantages of the dual aptamer architecture. While informative, the vast majority of these investigations was limited to *in vitro* techniques and often yielded conflicting results. In this study, we examined the proposed models of tandem glycine riboswitch regulation in the context of

its native locus within the *B. subtilis* genome and assessed how previously described interactions contribute to riboswitch function *in vivo* and overall cell fitness.

We find that mutations to the first aptamer (M1, M3, M5) resulted in the greatest reduction in downstream gene expression. All of these mutations directly or indirectly affect the structural stability and/or ligand-binding affinity of the first aptamer (Mandal et al. 2004; Kwon and Strobel 2008; Erion and Strobel 2011; Butler et al. 2011; Kladwang et al. 2011; Sherman et al. 2012; Baird and Ferré-D'Amaré 2013; Esquiaqui et al. 2014; Ruff and Strobel 2014). The M1 mutation to the glycine-binding domain on the first aptamer retained regulatory activity in response to glycine, although induced downstream expression was about one-tenth of that observed with the wild-type riboswitch. This suggests that glycine solely binding to the second aptamer can yield a regulatory response, but that glycine binding to the first aptamer is required for robust downstream gene expression. The first aptamer dimerization (M3) and leader-linker kink-turn (M5) mutations both completely abolished regulation and *gcvT* operon expression. This supports recent *in vitro* findings suggesting that aptamer dimerization is energetically linked to glycine binding in the first aptamer and that the ligand-binding affinity of the first aptamer is more sensitive to disruptions to the dimerization interface. These results also corroborate the importance of the P0/P1 helices for riboswitch dimerization and pre-organization of the ligand-binding pockets (Sherman et al. 2012; Ruff and Strobel 2014).

Mutations to the second aptamer had varying effects. Surprisingly, the M2 mutation to the glycine-binding domain on the second aptamer retained regulatory activity and expression levels comparable to that of the wild-type riboswitch. This reinforces the importance of glycine binding to the first aptamer for proper glycine riboswitch regulation and maximum downstream gene expression. Similarly, the M4

mutation to the dimerization domain on the second aptamer also retained regulatory function in response to glycine, although the resulting downstream gene expression was slightly less than that of the wild-type riboswitch. This finding supports previous studies suggesting that the second aptamer is capable of binding glycine despite disruptions to the dimerization interface (Ruff and Strobel 2014). While differences in behavior between the M3 and M4 dimerization mutants may be due to the asymmetrical nature of the γ dimerization interface (different nucleotide mutations are required to disrupt the interaction for each aptamer), our other mutations corroborate a model where changes to the first aptamer have a bigger impact on gene expression.

Taken together, our results indicate that double ligand occupancy is not required and glycine can bind to either aptamer to elicit a regulatory response *in vivo*. However, glycine binding to the first aptamer in combination with the leader-linker kink-turn and inter-aptamer interactions is necessary for optimal tandem glycine riboswitch-mediated regulation, maximum *gcvT* operon expression, and subsequent *B. subtilis* survival in high glycine environments. These findings offer new physiological insights into the decade-old dual aptamer debate and reinforce the hypothesis that the tandem architecture has been conserved against evolution not necessarily for enhanced ligand specificity and/or more digital gene control, but because the complex tertiary structure mediated by the presence of and interactions between the two aptamers is important for promoting glycine binding and conformational changes of the expression platform (Sherman et al. 2012; Ruff and Strobel 2014; Ruff et al. 2016). The recently proposed models of singlet glycine riboswitch regulation further support this theory. These single aptamer riboswitches bind glycine with affinities comparable to those with the dual aptamer structure, however, they are not true structural “singlets.” Interactions with stem-loop “ghost aptamers” are required for singlet riboswitch structural stability and

regulation, although these interactions may be less complex and/or distinct from those found with the tandem aptamer architecture (Ruff et al. 2016).

While we demonstrate that glycine binding to the second aptamer is not necessary to elicit the maximum regulatory response *in vivo*, previous bioinformatic analyses indicate that the glycine-binding domains on both aptamers are equally well-conserved (Barrick and Breaker 2007). This suggests that maintaining two functional ligand-binding domains may confer some evolutionary advantage over the singlet aptamer conformation. Despite the fact that our study does not directly suggest a cooperative mechanism of ligand binding, as this behavior is difficult to discern *in vivo*, the possibility that the tandem glycine riboswitch acts cooperatively within the cell cannot be precluded.

Although our overall conclusions are in agreement with the current model for tandem glycine riboswitch function, some of our data do not directly align with those from previous reports. Namely, we find that disruptions to the glycine-binding pocket on the second aptamer (M2) do not compromise riboswitch function or downstream expression *in vivo*, whereas *in vitro* studies with the *V. cholerae* riboswitch report that homologous mutations to the second aptamer have a more severe impact on tandem riboswitch ligand-binding affinity (Ruff and Strobel 2014). A similar trend is observed with our dimerization mutations. Our data suggest that mutations to the γ dimerization interface on the first aptamer (M3) completely abrogate riboswitch function, while mutations to the dimerization interface on the second aptamer (M4) still retain regulatory activity, albeit with reduced downstream expression. Past work with homologous *V. cholerae* riboswitch mutant constructs demonstrate that disrupting the γ dimerization domain on the second aptamer results in the greatest reduction in dimerization affinity between the two aptamers (Ruff and Strobel 2014).

The discrepancies between our results and those obtained with the *V. cholerae* riboswitch are most likely due to differences between *in vitro* and *in vivo* experimental conditions. Our experiments utilize the full-length *B. subtilis* tandem glycine riboswitch and our conclusions are based on *in vivo* gene expression and reporter enzyme activity levels, whereas previous investigations directly measured glycine-binding and dimerization affinities *in vitro* using both *cis* (full-length) and *trans V. cholerae* tandem glycine riboswitch constructs (Ruff and Strobel 2014). Additionally, though the tandem glycine riboswitches found in *B. subtilis* and *V. cholerae* are homologous, they are not identical and thus may not behave in the same manner in both *in vivo* and *in vitro* environments. For example, the *B. subtilis* tandem glycine riboswitch is followed by a Rho-independent terminator and its regulatory mechanism of action is well-characterized; the expression platform and mechanism by which the *V. cholerae* tandem glycine riboswitch regulates gene expression *in vivo* has not been characterized in detail (Mandal et al. 2004). Consequently, while general conclusions can be drawn across studies, direct comparisons may not be applicable.

Using *B. subtilis* as a model organism, we validate the current model of tandem glycine riboswitch folding *in vivo* and demonstrate the importance of proper tandem glycine riboswitch function as a genetic “on” switch for *gcvT* operon expression and optimal cell growth in both the presence and absence of glycine. The effects subtle point mutations to the tandem glycine riboswitch have on regulation, gene expression, and communal bacterial behaviors such as swarming and biofilm formation reinforce the potential of riboswitches as antimicrobial drug targets. Knowledge of how cells respond to the loss of riboswitch regulation allows for a more informed and directed approach to the design of riboswitch-targeting antibiotics and provides insight into how resistance to such compounds may evolve in the future.

4.4 Materials and Methods

4.4.1 β-galactosidase activity assay plasmid and strain construction

B. subtilis integration vector pDG1728 was modified to allow for a translational fusion between the glycine riboswitch constructs and *lacZ* (Guérout-Fleury et al. 1996). Briefly, the region containing the *spoVG* ribosome-binding site and the *lacZ* start codon was removed using the EcoRI and BamHI restriction sites and replaced with a cassette that allowed for a translational fusion with the *lacZ* reporter when cloning with BamHI:

5'-GAATTCTACGACAAATTGCAAAAATAATGTTGTCCTTTTAAATAAGATCTGATAAA ATGTGAACTAAGCTTCTAGGATCC-3'; underlining denotes EcoRI and BamHI restriction sites, respectively.

To generate the glycine riboswitch-*lacZ* translational fusion constructs, the region containing the *gcvT* operon promoter, wild-type glycine riboswitch, and ribosome-binding site and first nine codons of *gcvT* was PCR-amplified from *B. subtilis* 168 genomic DNA (GenBank: AL009126; complement of 2549307-2549704 (Kunst et al. 1997)) with primers containing EcoRI and BamHI restriction sites (Table 4.2). After digestion, the PCR product was cloned into the modified pDG1728 vector digested with the same enzymes. Mutations to the glycine riboswitch were obtained by site-directed mutagenesis or PCR assembly (Table 4.2). All plasmids were verified via Sanger sequencing.

Reporter constructs were transformed into *B. subtilis* 168 as described previously (Yasbin et al. 1975; Jarmer et al. 2002). Transformants were screened for resistance to spectinomycin (100 μg/mL), sensitivity to erythromycin (0.5 μg/mL), and loss of amylase activity (plating on TBAB + 1% starch and staining with Gram's iodine solution, Sigma-Aldrich) to ensure proper integration of the *lacZ* reporter constructs into the *amyE* locus.

4.4.2 β-galactosidase activity assays

B. subtilis 168 *lacZ* reporter strains were grown from single colonies in 2 mL M9 minimal medium + 1% glucose + 50 μg/mL tryptophan + 100 μg/mL spectinomycin for approximately 24 hours at 37ºC with shaking (225 rpm). These cultures were diluted 1:10 into 2 mL M9 minimal medium + 1% glucose + 50 μg/mL tryptophan + 100 μg/mL spectinomycin with varying glycine concentrations (0, 0.25, 0.5, and 1% w/v glycine, or approximately 0, 33.3, 66.6, and 133.2 mM glycine) and grown for approximately 10 hours at 37ºC with shaking (225 rpm). Cells (1.5 mL) were harvested and re-suspended in 1 mL Z buffer (50 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$, 50 mM 2-Mercaptoethanol) + 100 μg/mL spectinomycin. OD_{600} values were measured as 1:10 dilutions of cell suspensions in Z buffer. Samples with a final OD_{600} reading less than 0.5 were discarded. β-galactosidase activity assays were performed as previously described using 0.5 mL of cell suspensions and Miller Units were calculated as follows (Miller 1992):

Miller Units = $1000 * \frac{A_{420}}{\Delta t\ (\min) * A_{600} * vol.(mL)}$

The values reported represent three or more independent replicates; error bars represent standard error of the mean across biological replicates.

To assess the sensitivity of the M1 and M2 glycine-binding mutant riboswitch constructs relative to that of the wild-type glycine riboswitch, additional β-galactosidase activity assays were performed with these reporter strains as described above using the following glycine concentrations: 0.0025, 0.005, 0.01. 0.025, 0.05, and 0.1% w/v glycine, or approximately 0.33, 0.67, 1.33, 3.33, 6.66, and 13.32 mM glycine. The fraction of β-galactosidase activity for each riboswitch reporter construct at each glycine concentration tested was calculated as follows:

Fraction of activity = $\frac{Value\ (Miller\ Units) - Min.\ value\ (Miller\ Units\ at\ 0\%\ glycine)}{Max.\ value\ (Miller\ Units\ at\ 1\%\ glycine) - Min.\ value\ (Miller\ Units\ at\ 0\%\ glycine)}$

The values reported for these experiments represent two or more independent replicates; error bars represent standard error of the mean across biological replicates.

4.4.3 Native locus recombinant glycine riboswitch strain construction

To generate the recombinant strains for the growth curve, swarming motility, and biofilm assays, the *gcvT* operon promoter, wild-type glycine riboswitch, and two ~500 bp regions of homology flanking either side of the promoter and riboswitch region were PCR-amplified from *B. subtilis* 168 genomic DNA (GenBank: AL009126; complement of 2549705-2550291 for the 5' flanking ~500 bp region of homology, complement of 2549075-2549704 for the region containing the promoter, riboswitch, and 3' flanking ~500 bp region of homology). An erythromycin resistance cassette was PCR-amplified from *B. subtilis* integration vector pDG1663 (GenBank: U46200; 3930-5160) (Guérout-Fleury et al. 1996). PCR assembly with Phusion High-Fidelity DNA polymerase (ThermoFisher Scientific) was then used to generate recombinant PCR products in which the erythromycin resistance cassette was inserted immediately upstream from the *gcvT* operon promoter, in between the two regions of homology amplified from *B. subtilis* genomic DNA (Table 4.2). To generate the Δ*gcvT-gcvPB* construct, ~500 bp immediately downstream from the *gcvPB* coding region was PCR-amplified from *B. subtilis* 168 genomic DNA for the 3' region of homology (GenBank: AL009126; complement of 2544924-2545409). A double Rho-independent terminator construct was PCR-amplified from pYH213 (Yakhnin et al. 2015) (sequence the same as GenBank: AY599227; 631-768 (Choi et al. 2005)) and appended onto the 3' end of the erythromycin resistance cassette to prevent any read-through from the resistance cassette promoter. The complete Δ*gcvT-gcvPB* recombinant construct was then

assembled as described above. All assembly PCR constructs were poly-adenylated using Taq DNA polymerase (NEB), gel-purified, and cloned into pCR2.1 or pCR4 TOPO-TA vector (Invitrogen). Mutations to the glycine riboswitch were obtained by site-directed mutagenesis or PCR assembly (Table 4.2). All plasmids were verified via Sanger sequencing.

Recombinant pCR2.1 or pCR4 constructs were transformed into *B. subtilis* NCIB 3610 as described previously (Yasbin et al. 1975; Jarmer et al. 2002). Transformants were screened for resistance to erythromycin (0.5 μg/mL). Integration of the complete recombinant construct within the *gcvT* locus and the presence of the riboswitch mutations of interest were verified via PCR and Sanger sequencing.

4.4.4 Quantitative RT-PCR

Total RNA was extracted from early-to-mid log phase (OD_{600} ~0.4-0.6) *B. subtilis* NCIB 3610 strains grown in M9 minimal medium + 1% glucose +/- 0.25% glycine (+ 0.5 μg/mL erythromycin for recombinant strains) at 37ºC with shaking (225 rpm). To remove genomic DNA, 5 μg of total RNA was treated with RQ1 DNase (Promega) at 37ºC for 40 minutes followed by incubation at 98ºC for 2 minutes to heat-inactivate the enzyme, phenol-chloroform extraction, and ethanol precipitation. Reverse transcription was performed with random hexamers and SuperScript III according to the manufacturer's protocol (Invitrogen) and the resulting cDNA served as template for qPCR using an ABI 7500 Fast Real-Time PCR System (ThermoFisher Scientific) with SYBR Green detection. Glycine cleavage operon transcript levels were quantified using primers targeting the *gcvT* coding region and quantification of *nifU* was used as a normalization control (Table 4.2) (Reiter et al. 2011). To confirm effective removal of contaminating DNA, experiments were also conducted using reactions lacking reverse transcriptase.

Error bars represent standard error of the mean across three technical replicates propagated using standard calculations (Taylor 1997).

4.4.5 Growth curves with varying glycine concentrations

B. subtilis NCIB 3610 strains were grown from single colonies in 2 mL M9 minimal medium + 1% glucose (+ 0.5 μg/mL erythromycin for recombinant strains) for approximately 16 hours at 37ºC with shaking (225 rpm). These starter cultures were used to inoculate 0.5 mL M9 minimal medium + 1% glucose (+ 0.5 μg/mL erythromycin for recombinant strains) cultures with varying glycine concentrations (0, 0.25, 0.5, and 1% w/v glycine or approximately 0, 33.3, 66.6, and 133.2 mM glycine) in sterile non-treated 24-well cell culture plates to a starting OD_{600} reading of approximately 0.2. Plates were incubated at 37ºC with shaking (225 rpm) for 24 hours. OD_{600} values were recorded at one-hour intervals using a SpectraMax M3 Multi-Mode Microplate Reader (Molecular Devices). Doubling times were calculated as previously described, by taking the inverse of the slope of $\ln(OD_{600})$ in exponential phase readings (Rubinow 1975). The values reported represent three or more independent replicates; error bars represent the standard error of the mean across biological replicates. To determine the significance, mutant recombinant strain doubling times were compared to those of the wild-type recombinant strain at each glycine concentration using a Welch's single-tailed T-test in Microsoft Excel. Values were considered significantly different if $p<0.05$.

4.4.6 Growth assays with various carbon sources

B. subtilis NCIB 3610 starter cultures were prepared as described above and used to inoculate 0.5 mL M9 minimal medium +/- 0.25% glycine cultures with varying glucose concentrations (0, 0.2, 0.4, 0.6, 0.8, and 1% w/v glucose) in sterile non-treated

24-well cell culture plates to a starting OD_{600} reading of approximately 0.2. Plates were incubated at 37ºC with shaking (225 rpm) for approximately 24 hours. OD_{600} values were recorded at time points indicated using a SpectraMax M3 Multi-Mode Microplate Reader (Molecular Devices). For further analysis, select *B. subtilis* NCIB 3610 strains were patched onto the following plates (1.5% agar): M9 minimal medium + no carbon source, M9 minimal medium + 1% glucose, M9 minimal medium + 0.25% glycine, and M9 minimal medium + 1% glucose + 0.25% glycine. Plates were incubated at 37ºC for one week and photographs were taken using a Samsung WB380F digital camera.

4.4.7 Swarming motility assays

B. subtilis NCIB 3610 starter cultures were prepared as described above and grown to similar stationary phase OD_{600} values, and 5 μL of each culture was spotted onto the center of M9 minimal medium + 1% glucose + 0.7% agar swarming motility plates with the above glycine concentrations (Kearns and Losick 2003; Copeland and Weibel 2009). Plates were incubated at 37ºC for 48 hours and then photographed using a Samsung WB380F digital camera. The diameter of swarm motility growth was measured and recorded for each sample using FIJI software (Schindelin et al. 2012); three measurements were taken for each plate and averaged. The values reported represent three independent replicates; error bars represent standard error of the mean across biological replicates. To determine the significance, mutant recombinant strain swarm diameters were compared to those of the wild-type recombinant strain at each glycine concentration using a Welch's single-tailed T-test in Microsoft Excel. Values were considered significantly different if $p<0.05$. Representative photographs are shown.

4.4.8 Crystal violet staining

B. subtilis NCIB 3610 starter cultures were prepared as described above and grown to similar stationary phase OD_{600} values, and 1 μL of each starter culture was used to inoculate 100 μL MSgg minimal medium (5 mM potassium phosphate [pH 7], 100 mM MOPS [pH 7], 2 mM $MgCl_2$, 700 μM $CaCl_2$, 50 μM $MnCl_2$, 50 μM $FeCl_3$, 1 μM $ZnCl_2$, 2 μM thiamine, 0.5% glycerol, 0.5% glutamate, 50 μg/mL tryptophan, 50 μg/mL phenylalanine; + 0.5 μg/mL erythromycin for recombinant strains) cultures with the above glycine concentrations in sterile non-treated 96-well cell culture plates (Branda et al. 2001). Plates were incubated at 37ºC without agitation for 30 hours. Following incubation, culture supernatant was removed and discarded and the wells were washed twice with 100 μL 1X phosphate-buffered saline (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4). Plates were air-dried for 20 minutes and the remaining surface-attached cells were stained with 100 μL 0.1% Gram's crystal violet (BD Biosciences) for 20 minutes. Stain was then removed and plates were washed three times with 100 μL sterile water and allowed to air-dry for at least 30 minutes. Biofilm-associated crystal violet was then resolubilized in 100 μL 96% ethanol and incubated at room temperature for 10 minutes. For quantification, resolubilized samples were diluted 1:10 into 96% ethanol and OD_{570} measurements were recorded using a SpectraMax M3 Multi-Mode Microplate Reader (Molecular Devices). Wells incubated with cell-free medium were washed and stained as above to serve as a negative control and blank; the OD_{570} values from the cell-free wells were averaged and subtracted from the OD_{570} values for each sample (modified from (Kayumov et al. 2015)). The values reported represent six or more independent replicates; error bars represent standard error of the mean across biological replicates. To determine the significance, mutant recombinant strain crystal violet OD_{570} values were compared to those of the wild-type recombinant

strain at each glycine concentration using a Welch's single-tailed T-test in Microsoft Excel. Values were considered significantly different if $p<0.05$.

4.4.9 Pellicle assays

B. subtilis NCIB 3610 starter cultures were prepared as described above and grown to similar stationary phase OD_{600} values, and 1 μL of each culture was used to inoculate 0.5 mL MSgg minimal medium cultures (+ 0.5 μg/mL erythromycin for recombinant strains) with the above glycine concentrations in sterile non-treated 24-well cell culture plates (Branda et al. 2001). Plates were incubated at 37ºC without agitation for 40 hours. Wells were photographed at 24, 26, 28, 30, and 40 hours post-inoculation using a Samsung WB380F digital camera. This assay was repeated three independent times for each strain; representative photographs at 24 hours post-inoculation are shown.

4.4.10 Colony biofilm assays

B. subtilis NCIB 3610 starter cultures were prepared as described above and grown to similar stationary phase OD_{600} values, and 10 μL of each culture was spotted onto the center of MSgg minimal medium + 1.5% agar plates with the above glycine concentrations (Branda et al. 2001). Plates were incubated at 30ºC for 5 days and then photographed using a Samsung WB380F digital camera. This assay was repeated two independent times for each strain; representative photographs are shown.

Table 4.2. Oligonucleotides used in this study.
For each primer pair, the forward primer is listed first and the reverse primer is listed second.

Name	Sequence (5'-3')	Notes
1018	GGGCCC**GAATTC**CTGTTGCCCAGCATATAGTGATGATGGTA	Primer for cloning glycine riboswitch into modified pDG1728 reporter vector; EcoRI restriction site in bold
1019	CCGGCC**GGATCC**GTCAAATAACGGCGTTCTTTTCAGCAT	Primer for cloning glycine riboswitch into modified pDG1728 reporter vector; BamHI restriction site in bold
204	TATCTCTTGCCAGTCACGTTACG	Primer for PCR checks and sequencing of pDG1728 reporter constructs
122	GGGGACGACGACAGTATCGGCCTC	Primer for PCR checks and sequencing of pDG1728 reporter constructs
1014	TGACAAGATCATATGGGATAGACAG	Primer for amplifying 5'-500 bp region of glycine riboswitch recombinant construct
1015	CTTTAGGGTTATCGAATTCGATAAGCTTCTACAATTTGGGCAGATTTTTCTTATATTATTCATC	Primer for amplifying 5'-500 bp region of glycine riboswitch recombinant construct
1016	TAGCGCCTACGGGGAATTTGTATCGCGGCCGCCTGTTGCCCAGCATATAGTGATGATGGTAGG	Primer for amplifying 3'-500 bp region of glycine riboswitch recombinant construct
1017	TCGCTGTATATTGAGCACGGCCTGG	Primer for amplifying 3'-500 bp region of glycine riboswitch recombinant construct
681	TAGAAGCTTATCGAATTCGATAACCCTAAAG	Primer for amplifying erythromycin resistance cassette from pDG1663
682	GCGGCCGCGATACAAATTAAAAGTAGGCG	Primer for amplifying erythromycin resistance cassette from pDG1663
1046	CGAGCTTCCGGACAAATTCATAGTTC	Primer for confirming genomic integration of recombinant constructs
1047	GAAGCATTAATGACAAGCAGATAGCG	Primer for confirming genomic integration of recombinant constructs
745	GCAATGAAACACGCCAAAGTAAAC	Primer for PCR checks, sequencing recombinant glycine riboswitch constructs

1648	CTGCGGAGTGAATCTCACAGGCAAAAG AACTC	Mutagenesis primer for M1, glycine-binding mutation to first aptamer
1649	GAGTTCTTTTGCCTGTGAGATTCACTC CGCAG	Mutagenesis primer for M1, glycine-binding mutation to first aptamer
1650	GCAAAGTAAACTTACAGGTGCCAGGAC AGAG	Mutagenesis primer for M2, glycine-binding mutation to second aptamer
1651	CTCTGTCCTGGCACCTGTAAGTTTACTT TGC	Mutagenesis primer for M2, glycine-binding mutation to second aptamer
1263	GCAAACTGCGGAGTGAACCTCTCAGG CAAAAGAAC	Mutagenesis primer for M3, dimerization mutation to first aptamer
1264	GTTCTTTTGCCTGAGAGGTTCACTCCG CAGTTTGC	Mutagenesis primer for M3, dimerization mutation to first aptamer
1265	GCGTATGCAAAGTAAGCTTTCAGGTGC CAGG	Mutagenesis primer for M4, dimerization mutation to second aptamer
1266	CCTGGCACCTGAAAGCTTACTTTGCAT ACGC	Mutagenesis primer for M4, dimerization mutation to second aptamer
1526	CATGAAAATATGAGCGAATCCCAGCAA GGGGAGAGAC	Mutagenesis primer for M5, leader-linker kink-turn mutation
1527	GTCTCTCCCCTTGCTGGGATTCGCTCA TATTTTCATG	Mutagenesis primer for M5, leader-linker kink-turn mutation
1284	GGTGTTTCTCTGTAATTTTTTGTATG	Mutagenesis primer for M6, terminator mutation
1285	CATACAAAAAATTACAGAGAAACACC	Mutagenesis primer for M6, terminator mutation
1528	GACCTGACCGAAAATTTCGGGATACAG GCGC	Mutagenesis primer for M7, control mutation to first aptamer
1529	GCGCCTGTATCCCGAAATTTTCGGTCA GGTC	Mutagenesis primer for M7, control mutation to first aptamer
1530	GAGTGTTTGTGCGGAAGCGCAAACCAC CAAAGG	Mutagenesis primer for M8, control mutation to second aptamer
1531	CCTTTGGTGGTTTGCGCTTCCGCACAA ACACTC	Mutagenesis primer for M8, control mutation to second aptamer
1444	CTGACAGCTTCCAAGGAGCTAAAGAGG TCTCCTGTTGATAGATCCAGTAATGAC C	Primer for amplifying double terminator construct from pYH213 for appending onto 3' end of erythromycin resistance cassette from pDG1663 for Δ*gcvT-gcvPB* recombinant construct

1445	GGTCATTACTGGATCTATCAACAGGAG ACCTCTTTAGCTCCTTGGAAGCTGTCA G	Primer for amplifying double terminator construct from pYH213 for appending onto 3' end of erythromycin resistance cassette from pDG1663 for Δ*gcvT-gcvPB* recombinant construct
1446	GTTTAAACGATACAAATTCCCCGTAGG CGCTAGGGAAAAAAATTACGCCCCGCC CTGCC	Primer for appending double terminator construct onto 3' end of erythromycin resistance cassette from pDG1663 for Δ*gcvT-gcvPB* recombinant construct; use with primer 681
1652	CCCTAGCGCCTACGGGGAATTTGTATC GTTTAAACATAAAAACAGCTGTCTACCA GACAG	Primer for amplifying 3'-500 bp region of Δ*gcvT-gcvPB* recombinant construct
1653	CGAAAACGGCTCTATGACCTTG	Primer for amplifying 3'-500 bp region of Δ*gcvT-gcvPB* recombinant construct
1654	GAATCAGTTTATCAAACTGTCGGG	Primer for confirming genomic integration of Δ*gcvT-gcvPB* recombinant construct; use with primer 1046
1671	AAAGGAGAGAACCGCTATCTGC	Primer for qRT-PCR targeting the *gcvT* coding region
1672	AATCTGCACATCACCTGCTG	Primer for qRT-PCR targeting the *gcvT* coding region
1546	TTTTACTTCGTGACGGCGGT	Primer for qRT-PCR targeting *nifU*, the normalization control
1547	TTGTTGAACTTGGGCAGCTG	Primer for qRT-PCR targeting *nifU*, the normalization control

Chapter 5

Discussion

5.1 Summary and significance

This thesis offers novel insights into the *in vivo* behavior of RNA *cis*-regulators within the context of their native loci and highlights the advantages of combining computational, *in vitro*, and *in vivo* experimental approaches to obtain a more comprehensive understanding of regulatory RNA function from biochemical, biophysical, physiological, and evolutionary perspectives.

5.1.1 A conserved S6:S18-interacting RNA *cis*-regulator inhibits translation of *Escherichia coli rpsF*

In Chapter 2, I experimentally verify the *in vivo* regulatory activity of a conserved ribosomal protein-interacting mRNA structure that was computationally predicted by my group (and others) and previously shown to demonstrate robust interaction with the S6:S18 heterodimer *in vitro* (Matelska et al. 2013; Fu et al. 2014). I show that

overexpression of both S6 and S18 is required for regulation *in vivo*, I demonstrate that this RNA structure regulates gene expression at the translational level, and I confirm the predicted secondary structure and sites of RNA-protein interaction necessary for regulatory activity.

The S6:S18-interacting RNA *cis*-regulator presents a unique example of ribosomal protein autogenous regulation. The broad distribution of this RNA structure across many bacterial phyla is the exception, rather than the rule for such regulatory RNAs, as the majority of RNA *cis*-regulators identified to date are narrowly distributed to select subsets of bacteria. Furthermore, the S6:S18-interacting RNA regulates gene expression in response to interactions with a heterodimer composed of two secondary rRNA-binding proteins, whereas most known ribosomal protein RNA structures respond to interactions with a single primary rRNA-binding protein. The biological characterization of the S6:S18-interacting RNA allows us to fill in some of the existing gaps in our knowledge of ribosomal protein operon regulation in bacteria. Yet, it also reminds us that much remains to be discovered regarding this type of regulation, not only in non-model species of bacteria, but in the well-characterized model organism *E. coli* as well.

With that said, the work presented in Chapter 2 and the Appendix of this thesis emphasizes the importance of experimentally validating computationally predicted RNA structures. Over the past decade, advancements in bioinformatics and high-throughput sequencing methods have lead to the identification of many putative structured RNA motifs, yet few have been assigned actual biological function. The confirmation of the regulatory activity of the S6:S18-interacting RNA (Chapter 2) and the ongoing validation efforts for the putative L21-interacting RNA (Appendix) demonstrate the power of using computational tools for structured RNA discovery. However, the lack of apparent

regulatory function of the predicted L19 RNA leader in *B. subtilis* (Appendix) reveals the limitations that still exist with such *in silico* approaches. Making a more concerted effort to combine computational discovery with experimental validation will not only promote the development of improved prediction tools and subsequently expand existing RNA structural databases, but it will allow for more accurate discovery of novel, biologically relevant regulatory RNA structures. Structural and biophysical characterization of new regulatory RNA motifs will enable us to better compare RNA structures across diverse organisms to ultimately gain insight into the evolution of regulatory RNA structures in bacteria and how they co-evolve with their protein-binding partners.

5.1.2 Fitness advantages conferred by the L20-interacting RNA *cis*-regulator of ribosomal protein synthesis in *Bacillus subtilis*

The work presented in Chapter 3 takes the functional characterization of ribosomal protein RNA *cis*-regulators a step beyond the standard *in vitro* and *in vivo* validation techniques detailed in Chapters 1 and 2 to explore regulatory RNA behavior in its native context within the cell. It is well understood that these *cis*-regulatory RNA structures help ensure the stoichiometric production of ribosomal proteins for proper ribosome assembly. However, relatively little work has examined how the loss of RNA-mediated regulation of ribosomal protein operons affects ribosome assembly and cell fitness. In this Chapter, I use information from previous structural and functional studies to generate recombinant *B. subtilis* strains that harbor mutations designed to disrupt the regulatory activity of the L20-interacting RNA *cis*-regulator of the *infC* ribosomal protein operon. I examine how different mutations impact the regulation and expression of the *infC* operon within the recombinant strains and demonstrate that compromising the regulatory activity of the L20-interacting RNA causes cold temperature-sensitive defects

in growth, rRNA processing, and ribosomal subunit sedimentation. I find that point mutations to the L20-interacting RNA structure in the 5'-UTR of the *infC* ribosomal protein operon result in mutant phenotypes similar to those observed with bacterial strains carrying severe ribosomal protein mutations or deletions. My findings reinforce the sensitivity and cooperativity of ribosome assembly; even seemingly minute perturbations to the ribosome biosynthesis pathway can have detrimental effects on overall cell viability.

From an evolutionary perspective, this work improves our understanding of the forces that govern the evolution of RNA-based gene regulation in bacteria and the fitness advantages this type of regulation confers to bacteria. While this study examines only one example of a regulatory RNA structure that is narrowly distributed to Firmicutes, it is anticipated that most ribosomal protein RNA *cis*-regulators measurably contribute to bacterial fitness in some way. However, the extent of the fitness advantages these *cis*-regulatory RNAs confer to an organism may vary depending on the bacterial species, the RNA regulatory element itself, and/or the ribosomal protein operon it controls. Further analysis of both widely distributed and narrowly distributed examples of RNA *cis*-regulators in *B. subtilis* and other bacteria will allow us to better assess the evolutionary origins of and relationships between different RNA structures and the bacterial phyla in which they are found. In particular, such comparative studies will help elucidate whether the diverse regulatory structures we observe in nature originate from a distant common ancestor or are the result of multiple instances of independent evolution.

5.1.3 *In vivo* behavior of the tandem glycine riboswitch in *Bacillus subtilis*

In Chapter 4, I apply the same approach described in Chapter 3 to investigate the *in vivo* behavior of another, somewhat controversial, *cis*-regulatory RNA found in *B.*

subtilis: the tandem glycine riboswitch. Over the past decade, dozens of *in vitro* studies have examined the structural dynamics of the tandem glycine riboswitch in an effort to elucidate the evolutionary benefits of its unique dual aptamer composition. However, conflicting results stemming from differences in the conditions and constructs used for *in vitro* characterization have reopened the debate over the behavior and purpose of the tandem aptamer arrangement.

To shed some light on this controversy from a physiological perspective, I investigated the function and fitness contributions of the tandem glycine riboswitch within its native context in *B. subtilis*. I demonstrate that glycine binding to the first aptamer and first aptamer tertiary structure are required for proper glycine riboswitch function and maximum gene expression *in vivo* and that glycine riboswitch-mediated regulation is important for *B. subtilis* survival in high glycine environments. Although previous bioinformatics studies indicate that the ligand-binding domains on both aptamers are equally well conserved (Barrick and Breaker 2007), my results suggest that the conservation of the second aptamer may primarily be driven by the role its structure plays in mediating inter-aptamer contacts and conformational changes of the adjacent expression platform. The fact that glycine does not need to bind to the second aptamer to elicit the maximum regulatory response *in vivo* leads us to speculate that the ligand-binding activity of the second aptamer is potentially a vestigial function from the hypothesized duplication event that generated the tandem glycine riboswitch architecture (Ruff et al. 2016). Nevertheless, these hypotheses need to be examined in greater detail.

Overall, my conclusions biologically validate the *in vitro* folding model currently proposed for the tandem glycine riboswitch, however, some of my data differ from previous reports. These discrepancies are most likely due to differences between *in vitro*

and *in vivo* experimental methods and reinforce the necessity of characterizing riboswitch function and behavior using a combination of *in vitro* and *in vivo* approaches. My study is the first to examine the behavior of the tandem glycine riboswitch within its biological context, and while much remains to be elucidated regarding the origins, conservation, and selective advantages of its dual aptamer architecture, this project lays the necessary groundwork for such future investigations.

5.2 Broader impacts and future directions

5.2.1 Novel approach for studying RNA behavior and evolution *in vivo*

The methods employed in Chapters 3 and 4 offer a novel approach for investigating *cis*-regulatory RNAs within the context of their native loci. Apart from select regulatory assays (Chapter 2) and drug discovery efforts (see below), little work has examined how the results of *in vitro* and *in vivo* validation studies translate to actual physiological behavior of RNA *cis*-regulators within the cell or how the function of these regulatory RNA structures impact organism fitness. Using data obtained from previous characterization studies, I combine extensive mutational analyses of each candidate RNA, a knockout strategy in which the native locus of the regulatory RNA is replaced with a mutant version, and various phenotypic assays. This approach allows me to comprehensively examine the fitness costs associated with removing each RNA regulator as well as what RNA motifs are most essential for regulatory function and overall organismal viability. Examining regulatory RNA structure and function in this manner offers insight into the factors that drive RNA-ligand specificity and interactions, RNA structural diversity and conservation, and the selective advantages this type of RNA-based regulation confers to an organism. While mutant recombinant bacterial strains are widely used in the microbiology and molecular biology fields, most studies

utilize strains harboring gene deletions or mutations to protein coding regions. To my knowledge, I am the first to use native locus recombinant bacterial strains to assess how disruptions to regulatory RNA secondary structure impact gene regulation and cell fitness.

My approach also provides a suitable platform for studying the evolution of RNA structure *in vivo*, in real time. Although RNA is a common model "system" for evolutionary studies, most investigations are limited to *in vitro* selection methods, computational models, and RNA virus passaging studies. Very little work has explored how regulatory RNA structures actively emerge and co-evolve with their binding partners in a biological context. Compensatory evolution experiments conducted with RNA *cis*-regulator native locus mutant bacterial strains exhibiting significant fitness defects can improve our understanding of the balance required between sequence and structure conservation for biological function and how RNA structures explore sequence space *in vivo*. Moreover, the multiple distinct *cis*-regulatory RNA structures that interact with homologous ribosomal protein-binding partners (i.e. the S15- and L20-interacting RNAs) or the same small target molecule (i.e. the SAM, preQ1, c-di-GMP, guanine, and 2'-dG riboswitches) are excellent systems for examining how ligand specificity shapes RNA evolution *in vivo*, as the interactions between each unique RNA structure and its corresponding binding partner are often distinct for each RNA-ligand pair.

5.2.2 RNA *cis*-regulators as antimicrobial targets

Antibiotic resistance is emerging at an alarming rate within bacterial pathogens, thus limiting the use of once commonplace and clinically effective antibiotics (Andersson and Hughes 2010). Consequently, there is a growing need for the development of novel antimicrobial therapies that address new bacterial targets and use new modes of action,

while exhibiting little to no toxicity or cross-reactivity with targets within the patients they are meant to treat. Riboswitches and ribosomal protein RNA *cis*-regulators are both attractive targets for such new antibiotics, as they are mostly unique to bacteria, associated with genes essential for survival or pathogenesis, and interact with their binding partners with high specificity. Depending on their phylogenetic distribution, substances targeting certain *cis*-regulatory RNA structures can be used as broad-spectrum antibacterial drugs or as more selective antibiotics meant to target only specific bacterial phyla. My work reinforces this potential, as I show that removing *cis*-regulatory RNA-mediated gene regulation in *B. subtilis* results in severe defects in cell growth, ribosome assembly, and social behaviors such as swarming motility and biofilm formation; the latter two phenotypes in particular play key roles in protecting certain bacterial pathogens from antibiotics (Lai et al. 2009; Van Acker et al. 2014).

Riboswitches have been the focus of drug development efforts since their discovery almost 20 years ago and a handful of compounds have been identified and/or designed that interfere with riboswitch function and inhibit bacterial growth (for review see: (Blount and Breaker 2006; Deigan and Ferré-D'Amaré 2011)). An analog of thiamine, pyrithiamine, inhibits the growth of several bacterial and fungal species by targeting the TPP riboswitch and repressing genes involved in thiamine biosynthesis (Sudarsan et al. 2005). Two lysine analogs, L-aminoethylcysteine and DL-4-oxalysine, have been found to inhibit the growth of select Gram-positive bacteria through interfering with lysine riboswitch-mediated regulation of the lysine biosynthesis pathway (Blount et al. 2007). Roseoflavin, a riboflavin analog that is naturally produced by *Streptomyces davawensis*, and the synthetic FMN analog ribocil inhibit the growth of certain bacterial species by mechanisms involving the FMN riboswitch (Lee et al. 2009; Howe et al. 2015, 2016; Wang et al. 2017). Several compounds that bind guanine riboswitches and inhibit

bacterial growth have been developed (Kim et al. 2009; Mulhbacher et al. 2010) and the antibacterial activity of carba-α-D-glucosamine, a rationally designed GlcN6P analog that targets the *glmS* riboswitch, has recently been reported (Schüller et al. 2017).

In contrast, little to no research has explored the potential of ribosomal protein *cis*-regulatory RNAs as targets for novel antibiotics. However, given the similarities between these regulatory RNA structures and riboswitches, it is not difficult to apply the techniques used for riboswitch drug development to obtain antimicrobial compounds that interfere with the RNA-based regulation of ribosomal protein synthesis in bacteria. The prokaryotic ribosome is already a major antibiotic target and I, and others, have demonstrated that disruptions to ribosomal protein operon regulation, ribosomal protein overexpression, and/or alterations in ribosome composition have detrimental effects on bacterial fitness. Compounds that directly target or mimic these *cis*-regulatory RNA structures to either irreversibly repress ribosomal protein synthesis or allow for the de-regulation and subsequent uncontrolled synthesis of targeted ribosomal proteins can be developed. Such molecules would disrupt the stoichiometric production of ribosomal components required for proper ribosome assembly and ultimately inhibit bacterial growth.

My work assessing the *in vivo* behavior and fitness contributions of bacterial RNA *cis*-regulators provides a foundation for the future development of antimicrobials that target regulatory RNA structures. Knowledge of the fitness costs associated with the loss of RNA *cis*-regulator function as well as what RNA motifs are most essential for ligand binding and subsequent regulation *in vivo* allows for improved rational design of regulatory RNA-targeting compounds. Combining my approach with the evolution experiments described above enables us to investigate how bacteria evolve to

compensate for the loss of RNA *cis*-regulator-mediated gene control and informs us of how resistance to these novel antibiotics may eventually arise.

5.3 Concluding remarks

The computational discovery of structured RNA is a rapidly advancing field. Consequently, many regulatory RNA motifs have been predicted in bacterial genomes, yet relatively few have been biologically validated. Furthermore, a disconnect often exists between the data obtained from the *in vitro* and *in vivo* techniques used to characterize such RNAs; how an RNA structure folds and functions within a test tube or reporter construct does not necessarily reflect how it behaves within its biological context. This thesis reinforces the need to make a continued effort to experimentally verify the ever-growing pool of computationally predicted RNA *cis*-regulator candidates and demonstrates the benefits of examining *cis*-regulatory RNA behavior within its native context in the cell. I present an approach that not only allows for a more comprehensive characterization of RNA *cis*-regulator function *in vivo*, but also informs us about the selective forces that drive the emergence, evolution, and maintenance of structured RNA regulators in bacterial genomes. A greater understanding of the *in vivo* function and fitness contributions of different *cis*-regulatory RNAs in bacteria will significantly improve the existing models of structured RNA folding and evolution and aid the development of novel antimicrobials that target RNA regulators.

Appendix

Identification of putative RNA *cis*-regulator transcription start sites in *Bacillus subtilis*

A.1 Introduction

In 2007, Yao and colleagues published a computational tool for the *de novo* discovery of *cis*-regulatory RNAs in prokaryotic genomes. The pipeline integrates RNA motif prediction with RNA homology searches and genomic context. In a test with available sequenced Firmicute genomes, the method successfully recovered previously characterized regulatory RNA structures and predicted a number of additional RNA motifs that are promising candidates for novel regulatory elements. Of these candidates, three putative structured mRNA motifs preceding the genes encoding ribosomal proteins L13, L19, and L21 were reported in *B. subtilis* (Yao et al. 2007). Although this study was conducted almost a decade ago, the functions of these RNA elements remain

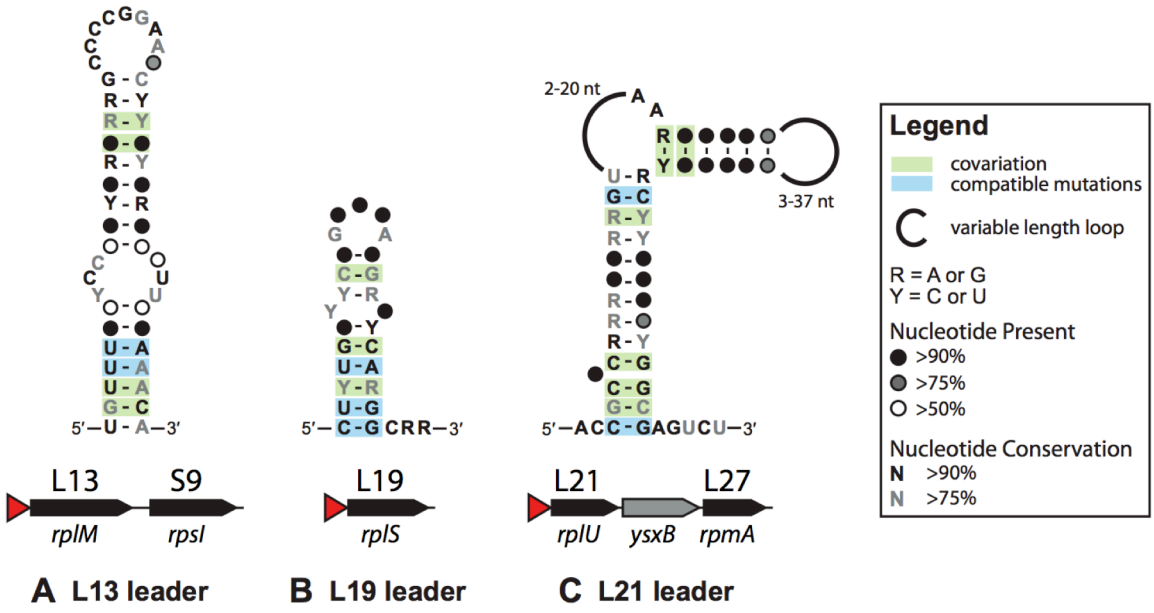


Figure A.1. Consensus sequences and secondary structures of the predicted *B. subtilis* ribosomal protein autoregulatory RNA motifs derived from (Yao et al. 2007). The genomic context of each putative RNA (indicated with a red arrow) is shown below its predicted structure. (A) The L13 RNA leader. (B) The L19 RNA leader. (C) The L21 RNA leader.

uncharacterized. Given the proximity of these predicted structures to ribosomal protein genes, it is hypothesized that these candidate RNAs are involved in the autogenous regulation of the ribosomal protein operons they precede. In an effort to validate the biological activity of the putative L13, L19, and L21 RNA leaders, we sought to confirm the transcription of each of these RNA elements in *B. subtilis* using 5'- and 3'-RACE and RT-PCR.

A.2 Results and Discussion

A.2.1 L13 leader

The L13 leader is a large bulged stem-loop that is predicted to precede the *rplM–rpsI* operon in *B. subtilis*, which encodes ribosomal proteins L13 and S9, respectively (Figure A.1A). Although L13 is a primary rRNA-binding protein that directly interacts with 23S rRNA during ribosome assembly, there is no homology between the L13-binding

site on the 23S rRNA and the predicted mRNA leader. Alternate pairings for this RNA structure that overlap the *rplM* start codon suggest a potential translational mechanism of regulation (Yao et al. 2007). Previous RNA-seq data confirms transcription of this putative RNA structure (Irnov et al. 2010) and we also verified transcription of the RNA leader region via RT-PCR. However, we were unable to recover transcription start sites for the *rplM-rpsI* operon using 5'-RACE with both log and stationary phase total RNA from *B. subtilis* 168. Additional RT-PCR analysis indicates that the predicted *rplM* promoter is co-transcribed with the putative L13 RNA leader and the *rplM* coding region, suggesting that the *rplM–rpsI* operon may be under the control of multiple promoters or that the 5'-UTR of this operon transcript may be longer than previously reported (Figure A.2A).

A.2.2 L19 leader

This putative RNA element is found upstream of the *rplS* gene, encoding ribosomal protein L19, which binds to the 23S rRNA at the interface between the 50S and 30S ribosomal subunits. This predicted structure consists of a small hairpin with an internal bulge and it is often found in close proximity to the Shine-Dalgarno sequence of the *rplS* open reading frame, indicating a potential translational regulatory mechanism (Figure A.1B). No sequence or structural similarities were identified between the putative regulatory mRNA structure and the 23S rRNA (Yao et al. 2007).

We recovered two transcription start sites for the *rplS* coding region: one approximately 7 nucleotides downstream from the predicted promoter sequence that includes the putative RNA leader region, and one immediately upstream from the *rplS* Shine-Dalgarno sequence (Figure A.2B). The regulatory potential of this RNA structure was assessed using β-galactosidase reporter assays similar to those described in

Chapter 3. Briefly, a construct containing the predicted L19 leader sequence originating from the 5'-most transcription start site recovered was cloned as a translational fusion with a *lacZ* reporter under the control of an IPTG-inducible promoter, and subsequently transformed into *B. subtilis* 168. No repression of reporter strain β-galactosidase activity was observed upon overexpression of ribosomal protein L19, suggesting that this putative RNA structure does not autogenously regulate *rplS* in response to L19 (Daniel Beringer, unpublished data). Finally, although *rplS* is predicted to be expressed as an independent transcriptional unit in Firmicutes, we were unable to confirm this with both RT-PCR and 3'-RACE (Yao et al. 2007).

A.2.3 L21 leader

The predicted L21 mRNA leader precedes the *rplU-ysxB-rpmA* operon (encoding ribosomal protein L21, hypothetical protein ysxB, and ribosomal protein L27, respectively) in *B. subtilis* and is composed of a large stem-loop with two highly conserved bulged adenosines (Figure A.1C). Both L21 and L27 are secondary rRNA-binding proteins during ribosome assembly; L21 interacts with L20 and the 23S rRNA, and L27 associates with L15 and the 5S rRNA. There is no apparent conservation between the putative mRNA structure and the 23S and 5S rRNAs (Yao et al. 2007).

Using 5'-RACE, we identified a transcription start site for the *rplU-ysxB-rpmA* operon approximately 8 nucleotides downstream from the predicted promoter sequence that contains the potential mRNA regulatory motif (Figure A.2C). We also confirmed that the L21 RNA leader, *rplU*, *ysxB*, and *rpmA* are co-transcribed in *B. subtilis* via RT-PCR (Figure A.2D). Moreover, β-galactosidase activity of a L21 RNA leader-*lacZ* translational fusion (as described above and in Chapter 3) is repressed approximately 2-fold upon overexpression of L21. No repression of β-galactosidase activity is observed in response

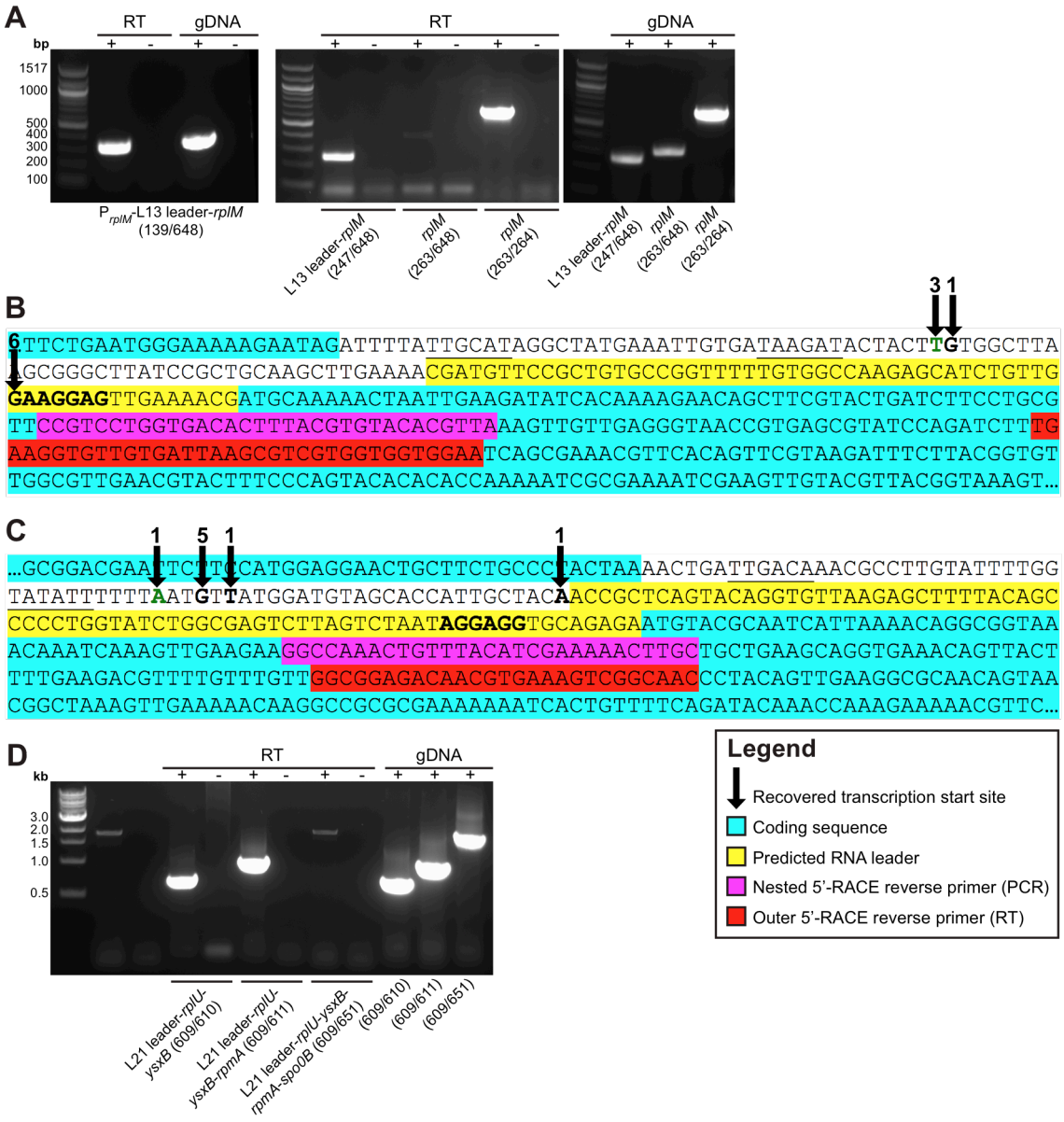


Figure A.2. Identification of predicted RNA leader transcription start sites in *B. subtilis* using 5'-RACE and RT-PCR. (A) Agarose gel electrophoresis (1%) of RT-PCR products from *B. subtilis* 168 log phase total RNA indicating that the predicted *rplM* promoter sequence, the putative L13 RNA leader, and *rplM* coding region are co-transcribed. (B) Transcription start sites recovered for the predicted L19 RNA leader. (C) Transcription start sites recovered for the predicted L21 RNA leader. Putative promoter sequences are underlined; putative Shine-Dalgarno sequences are in bold (Yao et al. 2007). The number of times each transcription start site was recovered is indicated above each arrow. The transcription start site used for subsequent analysis is in green. The sequences complementary to the reverse primers used for 5'-RACE are highlighted in magenta and red. (D) Agarose gel electrophoresis (1%) of RT-PCR products from *B. subtilis* 168 log phase total RNA demonstrating that the predicted L21 RNA leader, *rplU*, *ysxB*, and *rpmA* are co-transcribed.

to L27 overexpression. This suggests that the predicted L21 RNA leader is a regulatory element that inhibits downstream expression in response to L21 binding. Preliminary filter-binding assays confirm binding interactions between L21 and the RNA leader (Elizabeth Gray and Carolyn Larkins, unpublished data). Additional studies to further characterize the structure, sites of RNA-protein interaction, and regulatory mechanism of the L21 RNA motif are currently underway.

A.3 Materials and Methods

A.3.1 RT-PCR

B. subtilis 168 total RNA was extracted from log (OD_{600}~0.4-0.8) and stationary (OD_{600}~2.0+) phase cultures grown in 2XYT medium at 37ºC with shaking (225 rpm) as previously described. Genomic DNA was removed from 5 μg of total RNA by RQ1 DNase (Promega) digestion at 37ºC for 40 minutes, followed by heat-inactivation of the enzyme at 98ºC for 2 minutes, phenol-chloroform extraction, and ethanol precipitation. The DNase-treated RNA was reverse transcribed using SuperScript III and random hexamers according to the manufacturer's protocol (Invitrogen). PCR was conducted using this cDNA as template and gene-specific primers (Table A.1). To serve as controls, experiments were repeated with *B. subtilis* 168 genomic DNA and reverse transcription reactions lacking reverse transcriptase.

A.3.2 RACE

B. subtilis 168 total RNA was extracted from log and stationary phase cultures as described above and 5'-RLM-RACE was performed following the Invitrogen GeneRacer protocol with a homemade RNA-linker (Weinberg et al. 2009). First strands were synthesized using gene-specific primers and SuperScript III (Invitrogen) and the

reactions served as template for PCR using a forward primer corresponding to the 5'-linker and nested gene-specific reverse primers (Table A.1). PCR products were cloned into pCR2.1 TOPO-TA vector (Invitrogen) and sequenced to identify the 5' ends of the transcripts.

3'-RACE was conducted on log and stationary phase *B. subtilis* 168 total RNA according to the Invitrogen GeneRacer manual. *B. subtilis* 168 total RNA was polyadenylated and then reverse-transcribed using SuperScript III (Invitrogen) and a dT reverse primer. PCR was performed using the resulting cDNA as template with forward gene-specific primers and a nested dT reverse primer (Table A.1). PCR products were cloned into pCR2.1 TOPO-TA vector (Invitrogen) and sequenced to identify the 3' ends of the transcripts.

Table A.1. Oligonucleotides used in this study.
For primer pairs, the forward primer is listed first and the reverse primer is listed second.

Name	Sequence (5'-3')	Notes
8	TAATACGACTCACTATAGG	T7 promoter primer for generating 5'-RACE RNA-linker
7	TTTCTACTCCTTCAGTCCATGTCAGT GTCCTCGTGCTCCAGTCGCCTATAGT GAGTCGTATTA	Primer for generating 5'-RACE RNA-linker
9	GACTGGAGCACGAGGACACTGA	5'-RACE RNA-linker forward primer for PCR
604	GATCTCCAGTATCAACGTGTGGTGTG TAAGTTG	L13 RNA leader 5'-RACE outer reverse primer for RT
603	AGACGACCTAAAGTCTTGCCAGCAGC	L13 RNA leader 5'-RACE nested reverse primer for PCR; use with 9
650	GCCTGGATGTTGAGTGTGACGGTAG	L13 RNA leader 5'-RACE outer reverse primer 2 for RT
648	GTAAGTTGGTTTGTGTTTTCCGCGAA GG	L13 RNA leader 5'-RACE nested reverse primer 2 for PCR; use with 9
649	CCGCGAAGGATAGCTGCAACTTCTG	L13 RNA leader 5'-RACE nested reverse primer 3 for PCR; use with 9
291	CCCCAAGCTTATGGTATGTATTTCAA CCCCACGATAAGCCCCGG	L13 RNA leader forward primer for RT-PCR
247	CCCCAAGCTTATGGTATGTATTTCAA CCCCACGATAAG	L13 RNA leader forward primer 2 for RT-PCR
139	GGAATTCCTATGACAACTAAACCAGG TGTATAAT	*rplM* putative promoter forward primer for RT-PCR
263	GGACTAGTATGACCAAAGGAATCTTA GGAAGAAAAATTGG	*rplM* coding region forward primer for RT-PCR
264	CCGCGGATCCTTATTTAGATTTAACA GCACTTTTAACAGTG	*rplM* coding region reverse primer for RT-PCR
606	GTGCACGACGAGCGCCTTTAAGACC GTA	*rpsI* reverse primer for RT-PCR; use with 291 or 247
598	TTCCACCACCACGACGCTTAATCACA ACACCTTCA	L19 RNA leader 5'-RACE outer reverse primer for RT
597	TAACGTGTACACGTAAAGTGTCACCA GGACGG	L19 RNA leader 5'-RACE nested reverse primer for PCR; use with 9
599	ACGATGTTCCGCTGTGCCGGTTT	L19 RNA leader forward primer for RT-PCR
600	ATGACAGCCTGCCGAACTTTTCAGTG CG	*rbgA* reverse primer for RT-PCR; use with 599
601	GGAGCGAAAGTTTTGCGGTGCAATTC AGTTG	*rnhB* reverse primer for RT-PCR; use with 599

602	CTGTATTCGTCGGAGAAGGTGCTTTCGTA	*ylgG* reverse primer for RT-PCR; use with 599
10	GCGGTCACGCTTACTTAGCCCTCACTGAAATTTTTTTTTTTTTTTT	3'-RACE outer dT reverse primer for RT
674	CTTACGGTGTTGGCGTTGAACGTAC	L19 RNA leader 3'-RACE forward primer for PCR; use with 11
141	GGAATTCTAGATTTTATTGCATAGGCTATGAAATTGTGATA	*rplS* putative promoter 3'-RACE forward primer for PCR; use with 11
220	GACTAGTATGCAAAAACTAATTGAAGATATCAC	*rplS* coding region 3'-RACE forward primer for PCR; use with 11
11	GCGGTCACGCTTACTTAGCCCTCACTGAA	3'-RACE nested dT reverse primer for PCR
608	GTTGCCGACTTTCACGTTGTCTCCGCC	L21 RNA leader 5'-RACE outer reverse primer for RT
607	GCAAGTTTTTCGATGTAAACAGTTTGGCC	L21 RNA leader 5'-RACE nested reverse primer for PCR; use with 9
609	CCATTGCTACAACCGCTCAGTACAG	L21 RNA leader forward primer for RT-PCR
610	TAATCCCGTTCGATTGTCTCCAGCGAAAC	*ysxB* reverse primer for RT-PCR; use with 609
611	TTGCGGTCACGGCCGAAACGTTCGAAT	*rpmA* reverse primer for RT-PCR; use with 609
651	CGTGACTCGTGATTTCAAAACGCATG	*spo0B* reverse primer for RT-PCR; use with 609

References

Agalarov SC, Prasad G, Funke P, Stout C, Williamson JR. 2000. Structure of the S15, S6, S18-rRNA Complex: Assembly of the 30S Ribosome Central Domain. *Science* **288**: 107–112.

Ames TD, Breaker RR. 2011. Bacterial aptamers that selectively bind glutamine. *RNA Biol* **8**: 82–89.

Ames TD, Rodionov DA, Weinberg Z, Breaker RR. 2010. A Eubacterial Riboswitch Class That Senses the Coenzyme Tetrahydrofolate. *Chem Biol* **17**: 681–685.

Andersson DI, Hughes D. 2010. Antibiotic resistance and its cost: is it possible to reverse resistance? *Nat Rev Microbiol* **8**: 260–271.

Artsimovitch I, Henkin TM. 2009. In vitro approaches to analysis of transcription termination. *Methods* **47**: 37–43.

Aseev LV, Bylinkina NS, Boni IV. 2015. Regulation of the rplY gene encoding 5S rRNA binding protein L25 in Escherichia coli and related bacteria. *RNA* **21**: 851–861.

Aseev LV, Koledinskaya LS, Boni IV. 2016. Regulation of Ribosomal Protein Operons rplM-rpsI, rpmB-rpmG, and rplU-rpmA at the Transcriptional and Translational Levels. *J Bacteriol* **198**: 2494–2502.

Aseev LV, Levandovskaya AA, Tchufistova LS. 2008. A new regulatory circuit in ribosomal protein operons: S2-mediated control of the rpsB-tsf expression in vivo. *RNA* **14**: 1882–1894.

Babina AM, Lea NE, Meyer MM. 2017. In Vivo Behavior of the Tandem Glycine Riboswitch in Bacillus subtilis. *MBio* **8**: e01602-17.

Babina AM, Soo MW, Fu Y, Meyer MM. 2015. An S6:S18 complex inhibits translation of E. coli rpsF. *RNA* **21**: 1–8.

Baird NJ, Ferré-D'Amaré AR. 2013. Modulation of quaternary structure and enhancement of ligand binding by the K-turn of tandem glycine riboswitches. *RNA* **19**: 167–176.

Baker A, Draper DE. 1995. Messenger RNA Recognition by Fragments of Ribosomal Protein S4. *J Biol Chem* **270**: 22939–22945.

Baker JL, Sudarsan N, Weinberg Z, Roth A, Stockbridge RB, Breaker RR. 2012. Widespread Genetic Switches and Toxicity Resistance Proteins for Fluoride. *Science* **335**: 233–236.

Barrick JE, Breaker RR. 2007. The distributions, mechanisms, and structures of metabolite-binding riboswitches. *Genome Biol* **8**: R239.

Barrick JE, Corbino KA, Winkler WC, Nahvi A, Mandal M, Collins J, Lee M, Roth A, Sudarsan N, Jona I, et al. 2004. New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. *Proc Natl Acad Sci USA* **101**: 6421–6426.

Batey RT, Gilbert SD, Montange RK. 2004. Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature* **432**: 411–415.

Baughman G, Nomura M. 1983. Localization of the Target Site for Translational Regulation of the L11 Operon and Direct Evidence for Translational Coupling in Escherichia coli. *Cell* **34**: 979–988.

Benner SA, Ellington AD, Tauer A. 1989. Modern metabolism as a palimpsest of the RNA world. *Proc Natl Acad Sci USA* **86**: 7054–7058.

Bharat A, Brown ED. 2014. Phenotypic investigations of the depletion of EngA in Escherichia coli are consistent with a role in ribosome biogenesis. *FEMS Microbiol Lett* **353**: 26–32.

Blount KF, Breaker RR. 2006. Riboswitches as antibacterial drug targets. *Nat Biotechnol* **24**: 1558–1564.
Blount KF, Wang JX, Lim J, Sudarsan N, Breaker RR. 2007. Antibacterial lysine analogs that target lysine riboswitches. *Nat Chem Biol* **3**: 44–49.
Boni IV, Artamonova VS. 2000. The Last RNA-Binding Repeat of the Escherichia coli Ribosomal Protein S1 Is Specifically Involved in Autogenous Control. *J Bacteriol* **182**: 5872–5879.
Boni IV, Artamonova VS, Tzareva NV, Dreyfus M. 2001. Non-canonical mechanism for translational control in bacteria: synthesis of ribosomal protein S1. *EMBO J* **20**: 4222–4232.
Borovinskaya MA, Pai RD, Zhang W, Schuwirth BS, Holton JM, Hirokawa G, Kaji H, Kaji A, Cate JHD. 2007. Structural basis for aminoglycoside inhibition of bacterial ribosome recycling. *Nat Struct Mol Biol* **14**: 727–732.
Branda SS, González-Pastor JE, Ben-Yehuda S, Losick R, Kolter R. 2001. Fruiting body formation by Bacillus subtilis. *Proc Natl Acad Sci USA* **98**: 11621–11626.
Breaker RR. 2012. Riboswitches and the RNA world. *Cold Spring Harb Perspect Biol* **4**: 1–15.
Bremer H, Dennis PP. 1996. Modulation of Chemical Composition and Other Parameters of the Cell by Growth Rate. In *Escherichia coli and Salmonella* (ed. F.C. Neidhardt), pp. 1553–1569, ASM Press, Washington, DC.
Britton RA, Wen T, Schaefer L, Pellegrini O, Uicker WC, Mathy N, Tobin C, Daou R, Szyk J, Condon C. 2007. Maturation of the 5′ end of Bacillus subtilis 16S rRNA by the essential ribonuclease YkqC/RNase J1. *Mol Microbiol* **63**: 127–138.
Bruscella P, Shahbabian K, Laalami S, Putzer H. 2011. RNase Y is responsible for uncoupling the expression of translation factor IF3 from that of the ribosomal proteins L35 and L20 in Bacillus subtilis. *Mol Microbiol* **81**: 1526–1541.
Bucher T, Oppenheimer-Shaanan Y, Savidor A, Bloom-Ackermann Z, Kolodkin-Gal I. 2015. Disturbance of the bacterial cell wall specifically interferes with biofilm formation. *Environ Microbiol Rep* **7**: 990–1004.
Butler EB, Xiong Y, Wang J, Strobel SA. 2011. Structural basis of cooperative ligand binding by the glycine riboswitch. *Chem Biol* **18**: 293–298.
Cairns LS, Hobley L, Stanley-wall NR. 2014. Biofilm formation by Bacillus subtilis: new insights into regulatory strategies and assembly mechanisms. *Mol Microbiol* **93**: 587–598.
Cannone JJ, Subramanian S, Schnare MN, Collett JR, D'Souza LM, Du Y, Feng B, Lin N, Madabusi L V, Müller KM, et al. 2002. The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics* **3**: 1–31.
Castro-Roa D, Zenkin N. 2015. Methodology for the analysis of transcription and translation in transcription-coupled-to-translation systems in vitro. *Methods* **86**: 51–59.
Charollais J, Dreyfus M, Iost I. 2004. CsdA, a cold-shock RNA helicase from Escherichia coli, is involved in the biogenesis of 50S ribosomal subunit. *Nucleic Acids Res* **32**: 2751–2759.
Charollais J, Pflieger D, Vinh J, Dreyfus M, Iost I. 2003. The DEAD-box RNA helicase SrmB is involved in the assembly of 50S ribosomal subunits in Escherichia coli. *Mol Microbiol* **48**: 1253–1265.
Choi K-H, Gaynor JB, White KG, Lopez C, Bosio CM, Karkhoff-Schweizer RR, Schweizer HP. 2005. A Tn7-based broad-range bacterial cloning and expression system. *Nat Methods* **2**: 443–448.

Choonee N, Even S, Zig L, Putzer H. 2007. Ribosomal protein L20 controls expression of the Bacillus subtilis infC operon via a transcription attenuation mechanism. *Nucleic Acids Res* **35**: 1578–1588.
Choudhury P, Flower AM. 2015. Efficient assembly of ribosomes is inhibited by deletion of bipA in Escherichia coli. *J Bacteriol* **197**: 1819–1827.
Christensen T, Johnsen M, Fiil NP, Fniesen JD. 1984. RNA secondary structure and translation inhibition: analysis of mutants in the rplJ leader. *EMBO J* **3**: 1609–1612.
Copeland MF, Weibel DB. 2009. Bacterial swarming: a model system for studying dynamic self-assembly. *Soft Matter* **5**: 1174–1187.
Dai X, Zhu M, Warren M, Balakrishnan R, Patsalo V, Okano H, Williamson JR, Fredrick K, Wang YP, Hwa T. 2016. Reduction of translating ribosomes enables Escherichia coli to maintain elongation rates during slow growth. *Nat Microbiol* **2**: 16231.
Dambach M, Sandoval M, Waters LS, Dambach M, Sandoval M, Updegrove TB, Anantharaman V, Aravind L. 2015. The Ubiquitous yybP-ykoY Riboswitch Is a Manganese-Responsive Regulatory Element. *Mol Cell* **57**: 1099–1109.
Dean D, Nomura M. 1980. Feedback regulation of ribosomal protein gene expression in Escherichia coli. *Proc Natl Acad Sci USA* **77**: 3590–3594.
Deana A, Belasco JG. 2005. Lost in translation: The influence of ribosomes on bacterial mRNA decay. *Genes Dev* **19**: 2526–2533.
Deigan KE, Ferré-D'Amaré AR. 2011. Riboswitches: Discovery of Drugs that Target Bacterial Gene-Regulatory RNAs. *Acc Chem Res* **44**: 1329–1338.
Deiorio-Haggar K, Anthony J, Meyer MM. 2013. RNA structures regulating ribosomal protein biosynthesis in bacilli. *RNA Biol* **10**: 1180–1184.
Dubnau D. 1991. Genetic Competence in Bacillus subtilis. *Microbiol Rev* **55**: 395–424.
Erion TV, Strobel SA. 2011. Identification of a tertiary interaction important for cooperative ligand binding by the glycine riboswitch. *RNA* **17**: 74–84.
Esquiaqui JM, Sherman EM, Ionescu SA, Ye JD, Fanucci GE. 2014. Characterizing the dynamics of the leader-linker interaction in the glycine riboswitch with site-directed spin labeling. *Biochemistry* **53**: 3526–3528.
Fallon AM, Jinks CS, Strycharz GD, Nomura M. 1979. Regulation of ribosomal protein synthesis in Escherichia coli by selective mRNA inactivation. *Proc Natl Acad Sci USA* **76**: 3411–3415.
Fernández S, Ayora S, Alonso JC. 2000. Bacillus subtilis homologous recombination: Genes and products. *Res Microbiol* **151**: 481–486.
Feunteun J, Monier R, Vola C, Rosset R. 1974a. Ribosomal assembly defective mutants of Escherichia coli. *Nucleic Acids Res* **1**: 149–169.
Feunteun J, Rosset R, Ehresmann C, Stiegler P, Fellner P. 1974b. Abnormal maturation of precursor 16S RNA in a ribosomal assembly defective mutant of E . coli. *Nucleic Acids Res* **1**: 141–148.
Freedman LP, Zengel JM, Archer RH, Lindahl L. 1987. Autogenous control of the S10 ribosomal protein operon of Escherichia coli: genetic dissection of transcriptional and posttranscriptional regulation. *Proc Natl Acad Sci USA* **84**: 6516–6520.
Freeman WM, Walker J, Vrana KE. 1999. Quantitative RT-PCR: Pitfalls and Potential. *Biotechniques* **26**: 112–125.
Fu Y, Deiorio-Haggar K, Anthony J, Meyer MM. 2013. Most RNAs regulating ribosomal protein biosynthesis in Escherichia coli are narrowly distributed to Gammaproteobacteria. *Nucleic Acids Res* **41**: 3491–503.
Fu Y, Deiorio-Haggar K, Soo MW, Meyer MM. 2014. Bacterial RNA motif in the 5' UTR of rpsF interacts with an S6:S18 complex. *RNA* **20**: 168–176.
Furukawa K, Ramesh A, Zhou Z, Weinberg Z, Vallery T, Winkler WC, Breaker RR. 2015.

Bacterial Riboswitches Cooperatively Bind Ni2+ or Co2+ Ions and Control Expression of Heavy Metal Transporters. *Mol Cell* **57**: 1088–1098.
Gao A, Serganov A. 2014. Structural insights into recognition of c-di-AMP by the ydaO riboswitch. *Nat Chem Biol* **10**: 787–793.
Gayan M, Sherlock ME, Weinberg Z, Breaker RR. 2018. SAM-VI RNAs selectively bind S-adenosylmethionine and exhibit similarities to SAM-II riboswitches. *RNA Biol* [Epub ahead of print].
Gordon BJ, Hall RA, Stickland LH. 1951. The Kinetics of the Lysis of Bacterium Coli by Glycine. *J Hyg* **49**: 169–174.
Green NJ, Grundy FJ, Henkin TM. 2010. The T box mechanism: tRNA as a regulatory molecule. *FEBS Lett* **584**: 318–324.
Grundy FJ, Henkin TM. 1992. Characterization of the Bacillus subtilis rpsD Regulatory Target Site. *J Bacteriol* **174**: 6763–6770.
Grundy FJ, Henkin TM. 1991. The rpsD Gene, Encoding Ribosomal Protein S4, Is Autogenously Regulated in Bacillus subtilis. *J Bacteriol* **173**: 4595–4602.
Grundy FJ, Henkin TM. 1998. The S box regulon: a new global transcription termination control system for methionine and cysteine biosynthesis genes in Gram-positive bacteria. *Mol Microbiol* **30**: 737–749.
Guérout-Fleury A, Frandsen N, Stragier P. 1996. Plasmids for ectopic integration in Bacillus subtilis. *Gene* **180**: 57–61.
Guillier M, Allemand F, Dardel F, Royer CA, Springer M, Chiaruttini C. 2005a. Double molecular mimicry in Escherichia coli: binding of ribosomal protein L20 to its two sites in mRNA is similar to its binding to 23S rRNA. *Mol Microbiol* **56**: 1441–1456.
Guillier M, Allemand F, Graffe M, Raibaud S, Dardel F, Springer M, Chiaruttini C. 2005b. The N-terminal extension of Escherichia coli ribosomal protein L20 is important for ribosome assembly, but dispensable for translational feedback control. *RNA* **11**: 728–738.
Gusarov I, Nudler E. 1999. The Mechanism of Intrinsic Transcription Termination. *Mol Cell* **3**: 495–504.
Guthrie C, Nashimoto H, Nomura M. 1969. Structure and function of E. coli ribosomes: cold-sensitive mutants defective in ribosome assembly. *Proc Natl Acad Sci USA* **63**: 384–391.
Gyorfy Z, Draskovits G, Vernyik V, Blattner FF, Gaal T, Posfai G. 2015. Engineered ribosomal RNA operon copy-number variants of E. coli reveal the evolutionary trade-offs shaping rRNA operon number. *Nucleic Acids Res* **43**: 1783–1794.
Hall KB, Kranz J. 1999. Nitrocellulose Filter Binding for Determination of Dissociation Constants. In *RNA-Protein Interaction Protocols. Methods in Molecular Biology* (ed. S.R. Haynes), pp. 105–114, Humana Press, Totowa, NJ.
Hammes W, Schleifer KH, Kandeler O. 1973. Mode of Action of Glycine on the Biosynthesis of Peptidoglycan. *J Bacteriol* **116**: 1029–1053.
Harvey RJ. 1970. Regulation of Ribosomal Protein Synthesis in Escherichia coli. *J Bacteriol* **101**: 574–583.
Held WA, Ballou B, Mizushima S, Nomura M. 1974. Assembly Mapping of 30S Ribosomal Proteins from Escherichia coli. *J Biol Chem* **249**: 3103–3111.
Hellman LM, Fried MG. 2009. Electrophoretic Mobility Shift Assay (EMSA) for detecting protein-nucleic acid interactions. *Nat Protoc* **2**: 1849–1861.
Hishinuma F, Izaki K, Takahashi H. 1969. Effects of glycine and D-amino acids on growth of various microorganisms. *Agric Biol Chem* **33**: 1577–1586.
Hishinuma F, Izaki K, Takahashi H. 1971. Inhibition of L-Alanine Adding Enzyme by Glycine. *Agric Biol Chem* **35**: 2050–2058.

Howe JA, Wang H, Fischmann TO, Balibar CJ, Xiao L, Galgoci AM, Malinverni JC, Mayhood T, Villafania A, Nahvi A, et al. 2015. Selective small-molecule inhibition of an RNA structural element. *Nature* **526**: 672–677.

Howe JA, Xiao L, Fischmann TO, Wang H, Tang H, Villafania A, Zhang R, Barbieri CM, Roemer T. 2016. Atomic resolution mechanistic studies of ribocil: A highly selective unnatural ligand mimic of the E. coli FMN riboswitch. *RNA Biol* **13**: 946–954.

Huang L, Serganov A, Patel DJ. 2010. Structural Insights into Ligand Recognition by a Sensing Domain of the Cooperative Glycine Riboswitch. *Mol Cell* **40**: 774–786.

Iben JR, Draper DE. 2008. Specific interactions of the L10(L12)4 ribosomal protein complex with mRNA, rRNA, and L11. *Biochemistry* **47**: 2721–2731.

Irnov I, Sharma CM, Vogel J, Winkler WC. 2010. Identification of regulatory RNAs in Bacillus subtilis. *Nucleic Acids Res* **38**: 6637–6651.

Isono K, Cumberlidge AG, Isono S. 1977. Further temperature-sensitive mutants of Escherichia coli with altered ribosomal proteins. *Molec Gen Genet* **152**: 239–243.

Isono K, Kitakawa M. 1978. Cluster of ribosomal protein genes in Escherichia coli containing genes for proteins S6, S18, and L9. *Proc Natl Acad Sci USA* **75**: 6163–6167.

Isono K, Krauss J. 1976. Isolation and characterization of temperature-sensitive mutants of Escherichia coli with altered ribosomal proteins. *Molec Gen Genet* **149**: 297–302.

Jain C. 2008. The E. coli RhlE RNA helicase regulates the function of related RNA helicases during ribosome assembly. *RNA* **14**: 381–389.

Jarmer H, Berka R, Knudsen S, Saxild HH. 2002. Transcriptome analysis documents induced competence of Bacillus subtilis during nitrogen limiting conditions. *FEMS Microbiol Lett* **206**: 197–200.

Jinks-Robertson S, Nomura M. 1982. Ribosomal Protein S4 Acts in trans as a Translational Repressor to Regulate Expression of the alpha Operon in Escherichia coli. *J Bacteriol* **151**: 193–202.

Johnsen M, Christensen T, Dennis PP, Fiil NP. 1982. Autogenous control: ribosomal protein L10-L12 complex binds to the leader sequence of its mRNA. *EMBO J* **1**: 999–1004.

Kaczanowska M, Rydén-Aulin M. 2007. Ribosome biogenesis and the translation process in Escherichia coli. *Microbiol Mol Biol Rev* **71**: 477–94.

Kayumov A, Khakimullina E, Sharafutdinov I, Trizna E, Latypova L, Thi H, Margulis A, Bogachev M, Kurbangalieva A. 2015. Inhibition of biofilm formation in Bacillus subtilis by new halogenated furanones. *J Antibiot (Tokyo)* **68**: 297–301.

Kearns DB, Losick R. 2003. Swarming motility in undomesticated Bacillus subtilis. *Mol Microbiol* **49**: 581–590.

Kikuchi G, Motokawa Y, Yoshida T, Hiraga K. 2008. Glycine cleavage system: reaction mechanism, physiological significance, and hyperglycinemia. *Proc Jpn Acad Ser B* **84**: 246–263.

Kim JN, Blount KF, Puskarz I, Lim J, Link KH, Breaker RR. 2009. Design and antimicrobial action of purine analogues that bind guanine riboswitches. *ACS Chem Biol* **4**: 915–927.

Kim JN, Roth A, Breaker RR. 2007. Guanine riboswitch variants from Mesoplasma florum selectively recognize 2'-deoxyguanosine. *Proc Natl Acad Sci USA* **104**: 16092–16097.

Kim PB, Nelson JW, Breaker RR. 2015. An Ancient Riboswitch Class in Bacteria Regulates Purine Biosynthesis and One-Carbon Metabolism. *Mol Cell* **57**: 317–328.

Kladwang W, Chou F, Das R. 2011. Automated RNA structure prediction uncovers a missing link in double glycine riboswitches. *J Am Chem Soc* **134**: 1404–1407.

Knapp G. 1989. Enzymatic approaches to probing of RNA secondary structure and tertiary structure. *Methods Enzymol* **180**: 192–212.
Köhrer C, Mayer C, Neumair O, Gröbner P, Piendl W. 1998. Interaction of ribosomal L1 proteins from mesophilic and thermophilic Archaea and Bacteria with specific L1-binding sites on 23S rRNA and mRNA. *Eur J Biochem* **256**: 97–105.
Kunst F, Ogasawara N, Moszer I, Albertini AM, Alloni G, Azevedo V, Bertero MG, Bessières P, Bolotin A, Borchert S, et al. 1997. The complete genome sequence of the gram-positive bacterium Bacillus subtilis. *Nature* **390**: 249–256.
Kwasny SM, Opperman TJ. 2010. Static biofilm cultures of Gram-positive pathogens grown in a microtiter format used for anti-biofilm drug discovery. *Curr Protoc Pharmacol* **50**: 13A.8.1-13A.8.23.
Kwon M, Strobel SA. 2008. Chemical basis of glycine riboswitch cooperativity. *RNA* **14**: 25–34.
Lai S, Tremblay J, Déziel E. 2009. Swarming motility: a multicellular behaviour conferring antimicrobial resistance. *Environ Microbiol* **11**: 126–136.
Lee ER, Baker J., Weinberg Z, Sudarsan N, Breaker RR. 2010. An Allosteric Self-Splicing Ribozyme Triggered by a Bacterial Second Messenger. *Science* **329**: 845–849.
Lee ER, Blount KF, Breaker RR. 2009. Roseoflavin is a natural antibacterial compound that binds to FMN riboswitches and regulates gene expression. *RNA Biol* **6**: 1–8.
Lesage P, Chiaruttini C, Graffe M, Dondon J, Milet M, Springer M. 1992. Messenger RNA Secondary Structure and Translational Coupling in the Escherichia coli Operon Encoding Translation Initiation Factor IF3 and the Ribosomal Proteins, L35 and L20. *J Mol Biol* **228**: 366–386.
Lesage P, Truong H, Graffe M, Dondon J, Springer M. 1990. Translated Translational Operator in Escherichia coil: Auto-regulation in the infC-rpmI-rplT Operon. *J Mol Biol* **213**: 465–475.
Li G-W, Burkhardt D, Gross C, Weissman JS. 2014. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell* **157**: 624–635.
Li X, Yan Z, Xu J. 2003. Quantitative variation of biofilms among strains in natural populations of Candida albicans. *Microbiology* **149**: 353–362.
Lindahl L, Archer R, Zengel JM. 1983. Transcription of the S10 Ribosomal Protein Operon is Regulated by an Attenuator in the Leader. *Cell* **33**: 241–248.
Lindahl L, Zengel JM. 1979. Operon-specific regulation of ribosomal protein synthesis in Escherichia coli. *Proc Natl Acad Sci USA* **76**: 6542–6546.
Lindell M, Romby P, Wagner EGH. 2002. Lead (II) as a probe for investigating RNA structure in vivo. *RNA* **8**: 534–541.
Lipfert J, Das R, Chu VB, Kudaravalli M, Boyd N, Herschlag D, Doniach S. 2007. Structural Transitions and Thermodynamics of a Glycine-Dependent Riboswitch from Vibrio cholerae. *J Mol Biol* **365**: 1393–1406.
Lipfert J, Sim AYL, Herschlag D, Doniach S. 2010. Dissecting electrostatic screening, specific ion binding, and ligand binding in an energetic model for glycine riboswitch folding. *RNA* **16**: 708–719.
Luengpailin J, Doyle RJ. 2000. Glycine prevents the phenotypic expression of streptococcal glucan-binding lectin. *Biochim Biophys Acta* **1474**: 212–218.
Maciąg A, Peano C, Pietrelli A, Egli T, De Bellis G, Landini P. 2011. In vitro transcription profiling of the sigma-S subunit of bacterial RNA polymerase: Re-definition of the sigma-S regulon and identification of sigma-S-specific promoter sequence elements. *Nucleic Acids Res* **39**: 5338–5355.

Maguire BA. 2009. Inhibition of Bacterial Ribosome Assembly: a Suitable Drug Target? *Microbiol Mol Biol Rev* **73**: 22–35.
Mandal M, Breaker RR. 2004. Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nat Struct Mol Biol* **11**: 29–35.
Mandal M, Lee M, Barrick J, Weinberg Z, Emilsson G, Ruzzo W, Breaker R. 2004. A Glycine-Dependent Riboswitch That Uses Cooperative Binding to Control Gene Expression. *Science* **306**: 275–279.
Matelska D, Purta E, Panek S, Boniecki MJ, Bujnicki JM. 2013. S6:S18 ribosomal protein complex interacts with a structural motif present in its own mRNA. *RNA* **19**: 1341–1348.
McCown PJ, Corbino KA, Stav S, Sherlock ME, Breaker RR. 2017. Riboswitch diversity and distribution. *RNA* **23**: 995–1011.
McCown PJ, Liang JJ, Weinberg Z, Breaker RR. 2014. Structural, Functional, and Taxonomic Diversity of Three PreQ1 Riboswitch Classes. *Chem Biol* **21**: 880–889.
McGary K, Nudler E. 2013. RNA polymerase and the ribosome: the close relationship. *Curr Opin Microbiol* **16**: 112–117.
Meyer MM. 2017. The role of mRNA structure in bacterial translational regulation. *Wiley Interdiscip Rev RNA* **8**: 1–18.
Meyer MM, Ames TD, Smith DP, Weinberg Z, Schwalbach MS, Giovannoni SJ, Breaker RR. 2009. Identification of candidate structured RNAs in the marine organism "Candidatus Pelagibacter ubique." *BMC Genomics* **10**: 268.
Meyer MM, Roth A, Chervin SM, Garcia GA, Breaker RR. 2008. Confirmation of a second natural preQ1 aptamer class in Streptococcaceae bacteria. *RNA* **14**: 685–695.
Miller J. 1992. *A Short Course in Bacterial Genetics*. Cold Spring Harbor Laboratory Press.
Milligan JF, Groebe DR, Witherell GW, Uhlenbeck OC. 1987. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res* **15**: 8783–8798.
Mironov AS, Gusarov I, Rafikov R, Lopez LE, Shatalin K, Kreneva RA, Perumov DA, Nudler E. 2002. Sensing Small Molecules by Nascent RNA: A Mechanism to Control Transcription in Bacteria. *Cell* **111**: 747–756.
Mulhbacher J, Brouillette E, Allard M, Fortier L-C, Malouin F, Lafontaine DA. 2010. Novel Riboswitch Ligand Analogs as Selective Inhibitors of Guanine-Related Metabolic Pathways. *PLos Pathog* **6**: e1000865.
Muranaka N, Sharma V, Nomura Y, Yokobayashi Y. 2009. An efficient platform for genetic selection and screening of gene switches in Escherichia coli. *Nucleic Acids Res* **37**: e39.
Nahvi A, Barrick JE, Breaker RR. 2004. Coenzyme B12 riboswitches are widespread genetic control elements in prokaryotes. *Nucleic Acids Res* **32**: 143–150.
Nahvi A, Sudarsan N, Ebert MS, Zou X, Brown KL, Breaker RR, Haven N. 2002. Genetic Control by a Metabolite Binding mRNA. *Chem Biol* **9**: 1043–1049.
Nelson JW, Atilho RM, Sherlock ME, Stockbridge RB, Breaker RR, Nelson JW, Atilho RM, Sherlock ME, Stockbridge RB, Breaker RR. 2017. Metabolism of Free Guanidine in Bacteria Is Regulated by a Widespread Riboswitch Class. *Mol Cell* **65**: 220–230.
Nelson JW, Sudarsan N, Furukawa K, Weinberg Z, Wang JX, Breaker RR. 2013. Riboswitches in eubacteria sense the second messenger c-di-AMP. *Nat Chem Biol* **9**: 834–839.
Nelson JW, Sudarsan N, Phillips GE, Stav S, Lünse CE, McCown PJ, Breaker RR.

2015. Control of bacterial exoelectrogenesis by c-AMP-GMP. *Proc Natl Acad Sci USA* **112**: 5389–5394.
Nierhaus KH. 1991. The assembly of prokaryotic ribosomes. *Biochimie* **73**: 739–755.
Nomura M, Gourse R, Baughman G. 1984. The regulation of the synthesis of ribosomes and ribosomal components. *Annu Rev Biochem* **53**: 75–117.
Nomura M, Yates JL, Dean D, Post LE. 1980. Feedback regulation of ribosomal protein gene expression in Escherichia coli: Structural homology of ribosomal RNA and ribosomal protein mRNA. *Proc Natl Acad Sci USA* **77**: 7084–7088.
Nou X, Kadner RJ. 2000. Adenosylcobalamin inhibits ribosome binding to btuB RNA. *Proc Natl Acad Sci USA* **97**: 7190–7195.
Pan T. 2001. Probing RNA Structure by Lead Cleavage. In *Current Protocols in Nucleic Acid Chemistry*, p. 6.3.1-6.3.9.
Philippe C, Bénard L, Portier C, Westhof E, Ehresmann B, Ehresmann C. 1995. Molecular dissection of the pseudoknot governing the translational regulation of Escherichia coli ribosomal protein S15. *Nucleic Acids Res* **23**: 18–28.
Philippe C, Eyermann F, Bénard L, Portier C, Ehresmann B, Ehresmann C. 1993. Ribosomal protein S15 from Escherichia coli modulates its own translation by trapping the ribosome on the mRNA initiation loading site. *Proc Natl Acad Sci USA* **90**: 4394–4398.
Philippe C, Portier C, Mougell M, Ebell JP, Ehresmann B. 1990. Target Site of Escherichia coli Ribosomal Protein S15 on its Messenger RNA. *J Mol Biol* **211**: 415–426.
Poiata E, Meyer MM, Ames TD, Breaker RR. 2009. A variant riboswitch aptamer class for S-adenosylmethionine common in marine bacteria. *RNA* **15**: 2046–2056.
Portier C, Dondon L, Grunberg-Manago M. 1990. Translational Autocontrol of the Escherichia coli Ribosomal Protein S15. *J Mol Biol* **211**: 407–414.
Proshkin S, Rahmouni AR, Mironov A, Nudler E. 2010. Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science* **328**: 504–509.
Ramesh A, Winkler WC. 2017. Magnesium-sensing riboswitches in bacteria. *RNA Biol* **7**: 77–83.
Recht MI, Williamson JR. 2001. Thermodynamics and kinetics of central domain assembly. *J Mol Biol* **313**: 35–48.
Redko Y, Bechhofer DH, Condon C. 2008. Mini-III, an unusual member of the RNase III family of enzymes, catalyses 23S ribosomal RNA maturation in B. subtilis. *Mol Microbiol* **68**: 1096–1106.
Regulski EE, Breaker RR. 2008. In-line probing analysis of riboswitches. In *Methods in Molecular Biology* (ed. J. Wilusz), pp. 53–67, Humana Press, Totowa, NJ.
Regulski EE, Moy RH, Weinberg Z, Barrick JE, Yao Z, Ruzzo WL, Breaker RR. 2008. A widespread riboswitch candidate that controls bacterial genes involved in molybdenum cofactor and tungsten cofactor metabolism. *Mol Microbiol* **68**: 918–932.
Reiter L, Kolstø AB, Piehler AP. 2011. Reference genes for quantitative, reverse-transcription PCR in Bacillus cereus group strains throughout the bacterial life cycle. *J Microbiol Methods* **86**: 210–217.
Ren A, Patel DJ. 2014. c-di-AMP binds the ydaO riboswitch in two pseudo-symmetry-related pockets. *Nat Chem Biol* **10**: 780–787.
Rio DC. 2015. Northern blots: Capillary transfer of RNA from agarose gels and filter hybridization using standard stringency conditions. *Cold Spring Harb Protoc* 306–313.
Roth A, Winkler WC, Regulski EE, Lee BWK, Lim J, Jona I, Barrick JE, Ritwik A, Kim JN,

Welz R, et al. 2007. A riboswitch selective for the queuosine precursor preQ1 contains an unusually small aptamer domain. *Nat Struct Mol Biol* **14**: 308–317.
Rubinow S. 1975. *Introduction to Mathematical Biology*. John Wiley and Sons, New York, NY.
Ruff KM, Muhammad A, McCown PJ, Breaker RR, Strobel SA. 2016. Singlet glycine riboswitches bind ligand as well as tandem riboswitches. *RNA* **22**: 1–11.
Ruff KM, Strobel SA. 2014. Ligand binding by the tandem glycine riboswitch depends on aptamer dimerization but not double ligand occupancy. *RNA* **20**: 1775–1788.
Sandman K, Kroos L, Cutting S, Youngman P, Losick R. 1988. Identification of the promoter for a spore coat protein gene in Bacillus subtilis and studies on the regulation of its induction at a late stage of sporulation. *J Mol Biol* **200**: 461–473.
Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, et al. 2012. Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**: 676–682.
Schüller A, Matzner D, Lünse CE, Wittmann V, Schumacher C, Unsleber S, Brotz-Oesterhelt H, Mayer C, Bierbaum G, Mayer G. 2017. Activation of the glmS Ribozyme Confers Bacterial Growth Inhibition. *Chembiochem* **18**: 1–7.
Scott LG, Williamson JR. 2001. Interaction of the Bacillus stearothermophilus Ribosomal Protein S15 with its 5'-Translational Operator mRNA. *J Mol Biol* **314**: 413–422.
Scott M, Klumpp S, Mateescu EM, Hwa T. 2014. Emergence of robust growth laws from optimal regulation of ribosome synthesis. *Mol Syst Biol* **10**: 747.
Serganov A, Nudler E. 2013. A Decade of Riboswitches. *Cell* **152**: 17–24.
Serganov A, Polonskaia A, Ehresmann B, Ehresmann C, Patel DJ. 2003. Ribosomal protein S15 represses its own translation via adaptation of an rRNA-like fold within its mRNA. *EMBO J* **22**: 1898–1908.
Serganov A, Yuan Y-R, Pikovskaya O, Polonskaia A, Malinina L, Phan AT, Hobartner C, Micura R, Breaker RR, Patel DJ. 2004. Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. *Chem Biol* **11**: 1729–1741.
Sha Y, Lindahl L, Zengel JM. 1995. RNA Determinants Required for L4-mediated Attenuation Control of the S10 r-Protein Operon of Escherichia coli. *J Mol Biol* **245**: 486–498.
Shajani Z, Sykes MT, Williamson JR. 2011. Assembly of bacterial ribosomes. *Annu Rev Biochem* **80**: 501–26.
Shen P, Zengel JM, Lindahl L. 1988. Secondary structure of the leader transcript from the Escherichia coli S10 ribosomal protein operon. *Nucleic Acids Res* **16**: 8905–8924.
Sherlock ME, Breaker RR. 2017. Biochemical Validation of a Third Guanidine Riboswitch Class in Bacteria. *Biochemistry* **56**: 359–363.
Sherlock ME, Malkowski SN, Breaker RR. 2017. Biochemical Validation of a Second Guanidine Riboswitch Class in Bacteria. *Biochemistry* **56**: 352–358.
Sherman EM, Esquiaqui J, Elsayed G, Ye J-D. 2012. An energetically beneficial leader-linker interaction abolishes ligand-binding cooperativity in glycine riboswitches. *RNA* **18**: 496–507.
Slinger BL. 2016. Insights into the co-evolution of ribosomal protein S15 with its regulatory RNAs (Dissertation). Department of Biology, Boston College.
Slinger BL, Deiorio-Haggar K, Anthony JS, Gilligan MM, Meyer MM. 2014. Discovery and validation of novel and distinct RNA regulators for ribosomal protein S15 in diverse bacterial phyla. *BMC Genomics* **15**: 657.
Snell E, Guirard B. 1943. Some interrelationships of pyridoxine, alanine and glycine in

their effect of certain lactic acid bacteria. *Proc Natl Acad Sci USA* **29**: 66–73.
Srivastava AK, Schlessinger D. 1990. Mechanism and regulation of bacterial ribosomal RNA processing. *Annu Rev Microbiol* **44**: 105–129.
Sudarsan N, Cohen-Chalamish S, Nakamura S, Emilsson GM, Breaker RR, Haven N. 2005. Thiamine Pyrophosphate Riboswitches Are Targets for the Antimicrobial Compound Pyrithiamine. *Chem Biol* **12**: 1325–1335.
Sudarsan N, Lee ER, Weinberg Z, Moy RH, Kim JN, Link KH, Breaker RR. 2008. Riboswitches in Eubacteria Sense the Second Messenger Cyclic Di-GMP. *Science* **321**: 411–414.
Sudarsan N, Wickiser JK, Nakamura S, Ebert MS, Breaker RR. 2003. An mRNA structure in bacteria that controls gene expression by binding lysine. *Genes Dev* **17**: 2688–2697.
Sykes MT, Sperling E, Chen SS, Williamson JR. 2010. Quantitation of the ribosomal protein autoregulatory network using mass spectrometry. *Anal Chem* **82**: 5038–5045.
Tai PC, Kessler DP, Ingraham J. 1969. Cold-sensitive mutations in Salmonella typhimurium which affect ribosome synthesis. *J Bacteriol* **97**: 1298–1304.
Tang CK, Draper DE. 1989. Unusual mRNA pseudoknot structure is recognized by a protein translational repressor. *Cell* **57**: 531–536.
Taylor J. 1997. *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*. University Science Books, Sausalito, CA.
Tezuka T, Ohnishi Y. 2014. Two glycine riboswitches activate the glycine cleavage system essential for glycine detoxification in Streptomyces griseus. *J Bacteriol* **196**: 1369–1376.
Thomas MS, Bedwell DM, Nomural M. 1987. Regulation of Alpha Operon Gene Expression in Escherichia coli: A Novel Form of Translational Coupling. *J Mol Biol* **196**: 333–345.
Tissieres A, Watson JD, Schlessinger D, Hollingworth BR. 1959. Ribonucleoprotein particles from Escherichia coli. *J Mol Biol* **1**: 221–233.
Trausch JJ, Ceres P, Reyes FE, Batey RT. 2011. The Structure of a Tetrahydrofolate-Sensing Riboswitch Reveals Two Ligand Binding Sites in a Single Aptamer. *Structure* **19**: 1413–1423.
Van Acker H, Van Dijck P, Coenye T. 2014. Molecular mechanisms of antimicrobial tolerance and resistance in bacterial and fungal biofilms. *Trends Microbiol* **22**: 326–333.
Wang H, Mann PA, Xiao L, Flattery A, Mack M, Wang H, Mann PA, Xiao L, Gill C, Galgoci AM, et al. 2017. Dual-Targeting Small-Molecule Inhibitors of the Staphylococcus aureus FMN Riboswitch Disrupt Riboflavin Homeostasis in an Infectious Setting. *Cell Chem Biol* **24**: 576–588.
Wang JX, Breaker RR. 2008. Riboswitches that sense S-adenosylmethionine and S-adenosylhomocysteine. *Biochem Cell Biol* **86**: 157–168.
Wang JX, Lee ER, Morales DR, Lim J, Breaker RR. 2008. Riboswitches that Sense S-adenosylhomocysteine and Activate Genes Involved in Coenzyme Recycling. *Mol Cell* **29**: 691–702.
Weinberg Z, Barrick JE, Yao Z, Roth A, Kim JN, Gore J, Wang JX, Lee ER, Block KF, Sudarsan N, et al. 2007. Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline. *Nucleic Acids Res* **35**: 4809–4819.
Weinberg Z, Perreault J, Meyer MM, Breaker RR. 2009. Exceptional structured noncoding RNAs revealed by bacterial metagenome analysis. *Nature* **462**: 656–

659.
Weinberg Z, Wang JX, Bogue J, Yang J, Corbino K, Moy RH, Breaker RR. 2010. Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. *Genome Biol* **11**: R31.
Welz R, Breaker RR. 2007. Ligand binding and gene control characteristics of tandem riboswitches in Bacillus anthracis. *RNA* **13**: 573–582.
White HB. 1976. Coenzymes as Fossils of an Earlier Metabolic State. *J Mol Evol* **7**: 101–104.
Wilkinson KA, Merino EJ, Weeks KM. 2006. Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat Protoc* **1**: 1610–1616.
Winkler WC, Breaker RR. 2005. Regulation of Bacterial Gene Expression By Riboswitches. *Annu Rev Microbiol* **59**: 487–517.
Winkler WC, Cohen-Chalamish S, Breaker RR. 2002a. An mRNA structure that controls gene expression by binding FMN. *Proc Natl Acad Sci USA* **99**: 15908–15913.
Winkler WC, Nahvi A, Breaker RR. 2002b. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* **419**: 952–956.
Winkler WC, Nahvi A, Roth A, Collins JA, Breaker RR. 2004. Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* **428**: 281–286.
Winkler WC, Nahvi A, Sudarsan N, Barrick JE, Breaker RR. 2003. An mRNA structure that controls gene expression by binding S-adenosylmethionine. *Nat Struct Biol* **10**: 701–707.
Xu X, Ji Y, Stormo GD. 2009. Discovering cis-Regulatory RNAs in Shewanella Genomes by Support Vector Machines. *PLoS Comput Biol* **5**: e1000338.
Yakhnin H, Yakhnin AV, Babitzke P. 2015. Ribosomal protein L10(L12)4 autoregulates expression of the Bacillus subtilis rplJL operon by a transcription attenuation mechanism. *Nucleic Acids Res* **43**: 7032–7043.
Yao Z, Barrick J, Weinberg Z, Neph S, Breaker R, Tompa M. 2007. A Computational Pipeline for High-Throughput Discovery of cis-Regulatory Noncoding RNA in Prokaryotes. *PLoS Comput Biol* **3**: e126.
Yarnell WS, Roberts JW. 1999. Mechanism of Intrinsic Transcription Termination and Antitermination. *Science* **284**: 611–616.
Yasbin RE, Wilson GA, Young FE. 1975. Transformation and transfection in lysogenic strains of Bacillus subtilis: evidence for selective induction of prophage in competent cells. *J Bacteriol* **121**: 296–304.
Yates JL, Dean D, Strycharz WA, Nomura M. 1981. E. coli ribosomal protein L10 inhibits translation of L10 and L7/L12 mRNAs by acting at a single site. *Nature* **294**: 190–192.
Yates JL, Nomura M. 1980. E. coli ribosomal protein L4 is a feedback regulatory protein. *Cell* **21**: 517–522.
Zengel JM, Lindahl L. 1996. A Hairpin Structure Upstream of the Terminator Hairpin Required for Ribosomal Protein L4-Mediated Attenuation Control of the S10 Operon of Escherichia coli. *J Bacteriol* **178**: 2383–2387.
Zengel JM, Lindahl L. 1994. Diverse mechanisms for regulating ribosomal protein synthesis in Escherichia coli. *Prog Nucleic Acid Res Mol Biol* **47**: 331–343.
Zengel JM, Lindahl L. 1992. Ribosomal protein L4 and transcription factor NusA have separable roles in mediating termination of transcription within the leader of the S10 operon of Escherichia coli. *Genes Dev* **6**: 2655–2662.
Zengel JM, Sha Y, Lindahl L. 2002. Surprising flexibility of leader RNA determinants for r-protein L4-mediated transcription termination in the Escherichia coli S10 operon.

RNA **8**: 572–578.
Zhang J, Ferré-D'Amaré AR. 2015. Structure and mechanism of the T-box riboswitches. *Wiley Interdiscip Rev RNA* **6**: 419–433.
Zhou H, Zheng C, Su J, Chen B, Fu Y, Xie Y, Tang Q, Chou S-H, He J. 2016. Characterization of a natural triple-tandem c-di-GMP riboswitch and application of the riboswitch-based dual-fluorescence reporter. *Sci Rep* **6**: 20871.